17579

NATIONAL LIBRARY
OTTAWA

CANADA

BIBLIOTHÈQUE NATIONALE
OTTAWA

NAME OF AUTHOR......Michael James Kozlow

TITLE OF THESIS...Test Development For the
Measurement Of Scientific Attitude
As Inferred From Student's Expressed Behavioral Intent

UNIVERSITY...University of Alberta, Edmonton

DEGREE FOR WHICH THESIS WAS PRESENTED.....M. Ed.

YEAR THIS DEGREE GRANTED.........1973

Permission is hereby granted to THE NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

(Signed)...James Kozlow

PERMANENT ADDRESS:

17[illegible]5 - 78 Ave
Edmonton
Alberta

DATED...July 30.......1973

NL-91 (10-68)

THE UNIVERSITY OF ALBERTA

TEST DEVELOPMENT FOR THE MEASUREMENT OF SCIENTIFIC ATTITUDE AS INFERRED FROM STUDENTS' EXPRESSED BEHAVIORAL INTENT

by

Ⓒ MICHAEL JAMES KOZLOW

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE
OF MASTER OF EDUCATION

DEPARTMENT OF SECONDARY EDUCATION

EDMONTON, ALBERTA

FALL, 1973

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, the thesis entitled, "Test Development For The Measurement Of Scientific Attitude As Inferred From Students' Expressed Behavioral Intent", submitted by M. James Kozlow in partial fulfilment of the requirements for the degree of Master of Education.

M. A. Nay
Supervisor

H. Kass

A R Hakstian

Date May 10, 1973

Abstract

A rationale for evaluation in the affective domain of science education was developed. The procedures and data analysis employed were designed to demonstrate this rationale as a practical approach to test construction in this area.

The Nay-Crocker inventory of affective attributes of scientists was used as a framework for affective objectives in science education. A selected set of attributes (critical mindedness, suspended judgement, respect for evidence, honesty, objectivity, and willingness to change opinions) was behaviorially defined and multiple-choice questions to reflect these behaviors were constructed (Test On Scientific Attitude - TOSA).

Cognitive, intent, and action components of the attributes were defined and TOSA was divided into two subtests. The Cognitive Component Subtest (CCS) measures understanding of how the defining behaviors are manifest in the activities of scientists. The Intent Component Subtest (ICS) requires indication of a preference for a given course of action in situations related to the defining behaviors. Teacher ratings of student affective behavior were also obtained.

Item analysis results indicate that, while some of the items require revision, the statistics for most of the items

are satisfactory. The KR-20 coefficients (0.55, 0.45, and 0.39 for TOSA, CCS, and ICS, respectively) are quite low; however, the test-retest correlations (0.71, 0.68, and 0.64 for TOSA, CCS, and ICS, respectively) are satisfactory.

The item-intercorrelations for TOSA were examined by a common factor analysis and nine factors were retained. When each factor was identified with one of the attributes, 80% of the salient factor loadings were related to a classification of the questions based on the defining behaviors. The factor solution gave some support to the contention that CCS and ICS measure different characteristics. Four of the factors consist mainly of questions from ICS and two consist mainly of questions from CCS.

This division into two subtests is also supported by a number of correlations. The correlation between CCS and ICS is 0.23 again indicating that the two subtests do not measure the same characteristics. ICS is more highly correlated with teacher ratings of student affective behavior while CCS is more highly correlated with scholastic ability and reading ability.

A test consisting of opinion statements for a Likert scale (TOLI) was designed to provide a comparison between this format and the format of TOSA. Although statements were included in TOLI only if their content was similar to

the content of TOSA, the correlation between these two tests is only 0.37 indicating that test format as well as content may have some influence on student responses.

High and low student groups as categorized by teacher ratings were shown to be significantly different ($p \leq 0.001$) by one-way analysis of covariance with scholastic ability as the covariate and TOSA, CCS, ICS, and TOLI as the separate criterion measures.

Although weaknesses were identified in some of the questions, the data analysis indicates that useful tests can be constructed through the application of the rationale outlined in this study.

ACKNOWLEDGEMENTS

My sincere thanks go to my supervisor, Dr. Marshall Nay, for his valuable advice and for his continued support and encouragement. I must also acknowledge his assistance in the construction of the test questions.

I also appreciate the assistance provided by Dr. Heidi Kass who acted as my supervisor in Dr. Nay's absence and continued to provide valuable advice at all stages of the development of this study.

Thanks also go to Dr. T. Maguire and Dr. R. Hakstian for their assistance in selecting appropriate data analysis procedures and in interpreting the results.

I wish to thank the students and staff of Jasper Place Composite High School, M.E. LaZerte Composite High School, and Harry Ainlay Composite High School for their participation.

I wish to thank the students and staff in the Department of Secondary Education at The University of Alberta for their assistance in the selection of behavioral objectives.

I wish to thank the Division of Educational Research Services for the use of their computer programs.

I also wish to express my appreciation to my typist, Lana Ray.

TABLE OF CONTENTS

CHAPTER PAGE

I. THE PROBLEM 1

Need for the Study 1

Background of the Problem 2

Statement of the Problem 5

Definition of Terms 5

Questions and Hypotheses 7

Delimitations 10

Limitations 11

Experimetnal Procedures and Design 11

II. REVIEW OF THE RELATED LITERATURE 13

Defining the Affective Domain 13

Behavioral Objectives 19

Techniques Used in Attitude Measurement . . . 23

Summary of Research on Attitude Measurement in Science Education 28

III. EXPERIMENTAL PROCEDURES AND DESIGN 36

A Rationale for the Construction Of the Test Items 36

Data Collection 48

Data Analysis 55

IV. RESULTS AND DISCUSSION 63

Content Validity 64

Descriptive Statistics 69

Structural Validity 77

Test Stability 100

External Validty 102

CHAPTER PAGE

V. SUMMARY AND RECOMMENDATIONS 113
  Summary 113
  Implications for Classroom Evaluation . . . 119
  Recommendations 120

* * *

BIBLIOGRAPHY 125

APPENDIX A. AFFECTIVE ATTRIBUTES OF SCIENTISTS . . . 131

APPENDIX B. TEST ON SCIENTIFIC ATTITUDE 134

APPENDIX C. TEST ON LIKERT ITEMS 147

APPENDIX D. INSTRUCTIONS TO PANEL FOR RATING BEHAVIORS 151

APPENDIX E. INSTRUCTIONS TO TEACHERS FOR STUDENT RATINGS 156

APPENDIX F. CORRELATION MATRIX FOR TEST ITEMS . . . 159

APPENDIX G. INTERCORRELATIONS FOR THE NINE FACTORS . 161

LIST OF TABLES

Table	Description	Page
I	F Tests for Variance Differences Between GROUP 1 and GROUP 2	70
II	T Tests for Mean Differences Between GROUP 1 and GROUP 2	70
III	Test Means, Standard Deviations, and Ranges	72
IV	Frequency Distributions of Test Scores	73
V	Correlations for the Sample of 307 Students	73
VI	Correlations for GROUP 1	74
VII	Analysis of Covariance on TOSA Scores - Males vs Females	76
VIII	Analysis of Covariance on CCS Scores - Males vs Females	76
IX	Analysis of Covariance on ICS Scores - Males vs Females	77
X	Difficulty Levels and Biserial Correlations For the Test Items in TOSA	80
XI	Item to Total Correlations for TOLI	83
XII	Oblique Factor Pattern Matrix for Nine Factors of the Thirty-nine Test Items in TOSA	87
XIII	Test-retest Correlations for 39 Test Items In TOSA for 105 Students	101
XIV	Means and Standard Deviations For Test-retest Data on TOSA, CCS, and ICS for 105 Students	102
XV	Analysis of Covariance on TOSA Scores Low Student Group vs High Student Group	104
XVI	Analysis of Covariance on CCS Scores - Low Student Group vs HIGH Student Group	105
XVII	Analysis of Covariance on ICS Scores Low Student Group vs High Student Group	105
XVIII	Analysis of Covariance on TOLI Scores - Low Student Group vs High Student Group	106

TABLE	DESCRIPTION	PAGE
XIX	High-Low Group Comparisons for Each Test Item in TOSA	109
XX	Correlations of Test Scores With Teacher ratings	111

LIST OF FIGURES

Figure | | Page
I A Model for Educational Planning 37

Chapter I

The Problem

I. Need for the Study

Lists of objectives proposed for science education often include the development of interests, attitudes, values, and appreciations (e. g. Alberta Department Of Education, 1970). However, the literature on this topic (Bloom, et. al., 1971, p. 226; Nay and Crocker, 1970, p. 61) indicates that teachers tend to neglect these objectives when planning classroom activities.

In their discussion of summative and formative evaluation, Bloom and co-authors (1971, pp. 226-254) discuss possible reasons for this avoidance of direct attention to affective objectives. They claim that there is a general feeling, both among the public and the teaching profession, that trying to develop selected attitudes and values in students is akin to indoctrination and brainwashing. Scriven (1966, pp. 44-55) maintains that teaching toward affective objectives should be approached in a manner similar to the teaching of many science concepts. Selected values and attitudes should be presented as the most defensible ones from a given set of alternatives. Emphasis should be placed on developing an understanding of the arguments in support of those which are selected.

Another contributing factor to the neglect of affective objectives is the absence of adequate information concerning teaching methods and materials that can be used to develop interests, attitudes and values, and the lack of suitable evaluation instruments. This study will specifically investigate the problem of evaluation in the affective domain of science education. Research into the effectiveness of instructional methods in this area will be quite difficult unless appropriate evaluation instruments are available.

II. Background of the Problem

Any study involving evaluation must be concerned with the identification of the set of objectives which defines the area of interest, the detailed definition of these objectives, and the choice of an appropriate format for the evaluation instruments. As these points are examined, the function that the evaluation instruments are to serve will be a dominant factor affecting choices between various alternatives.

The main purpose of evaluation in the affective domain should be to guide the development and improvement of teaching methods and materials rather than to assign pupil marks (Bloom, et. al., 1971, p. 228). A possible approach to meet this end is to construct evaluation instruments which attempt to identify behaviors that define the selected set of objectives. Student scores can then be used as an indication of the extent to which the applied methods and materials encourage the desired behaviors. The preparation of these instruments will be facilitated if the objectives are defined in behavioral terms.

The affective domain in science has been defined in terms of affective attributes which scientists exhibit in their work and in their relationships with other scientists. These attributes are categorized as interest, adjustments, attitudes, appreciations and values (Nay and Crocker, 1970, pp. 61-62). Current lists of objectives for science education (Alberta Department of Education, 1970) include the development of many of the characteristics which are included in the summary presented by Nay and Crocker. If these characteristics are defined in terms of student behaviors, the Nay-Crocker inventory provides a comprehensive list of affective objectives for science education. The characteristics to be examined in the present study will be selected from this inventory. The complete inventory is listed in Appendix A.

The choice of test-question format to be used in the present study to measure student attitudes was largely influenced by the definiton of attitude provided by Rokeach (1968, p. 112). He defines attitude as a "relatively enduring organization of beliefs around an object or situation predisposing one to respond in some preferential manner".

Since attitudes are mediating variables, they cannot be measured directly and must be inferred from some overt response. The most common approach to attitude measurement is to obtain a measure of the respondent's agreement or disagreement with a set of opinion statements about the attitude object. The position taken in the present study is that, since attitudes are defined as predispositions to some preferred response, a reasonable approach to attitude measurement would be to make inferences about an individual's attitudes from his endorsement, or lack of it, of various courses of action in certain situations relevant to the attitude object.

Rokeach discusses three components of attitude - cognitive, affective,and behavioral. These three components represent knowledge about the attitude object, the tendency to take a positive or negative position toward the attitude object, and overt responses with respect to the attitude object. Nay and Crocker(1970, p. 61) define two components of the affective attributes in terms of student objectives. These are the student's cognition of the role of the attributes in the activities of scientists and the student's tendency to exhibit these attributes in his own science work. These correspond to the cognitive and behavioral components defined by Rokeach.

In the process of defining the affective attributes in behavioral terms, the present study incorporates the concepts presented by both Rokeach and Nay & Crocker to define three components of the affective attributes of scientists. These will be identified throughout this study as the cognitive, intent, and action components. The cognitive component represents the student's understanding of the significance of the attributes to the scientist in his work. The intent component represents the student's tendency to show approval or disapproval of behaviors which define the attribute. This will be indicated by his endorsement of specific courses of action in certain situations relevant to the attribute. The action component represents the extent to which the student demonstrates the attributes in his science work. The reasons for using these three terms are discussed in Chapter II.

III. Statement of the Problem

The problem in this study is to develop a rationale for evaluation of affective objectives in science education and to examine the practicality of this rationale through its application to a small subset of objectives. This will involve the definition of these objectives and the construction and field-testing of test items.

IV. Definition of Terms

Affective Objectives: For the purpose of this study the term affective objectives refers to the development of the interests, attitudes, adjustments, appreciations and values which are summarized in the Nay-Crocker inventory (see Appendix A).

Attitudes: An attitude is a "relatively enduring organization of beliefs around an object or situation predisposing one to respond in some preferential manner" (Rokeach, 1968, p. 112). In the present study, attitudes are further defined in terms of the following three components:

1. Cognitive Component - student's understanding of the significance of the affective attribute (Appendix A) in a scientist's work.

2. Intent component - student's tendency to show approval or disapproval of behaviors which define the affective attributes.

3. Action Component - student's tendency to exhibit the affective attributes in his science work.

Test On Scientific Attitude (TOSA): This is a forty-item, multiple-choice test developed as a part of this study (see Appendix B). The test items are divided into the following two subtests:

1. Cognitive Component Subtest (CCS) - The twenty items in this subtest are designed to measure the student's understanding of how the behaviors which define the affective attributes are manifest in the activities in which scientists participate.

2. Intent Component Subtest (ICS) - The twenty items in this subtest require the student to show a preference for a given course of action in a certain situation.

High Student Group: The teachers were asked to rate the students who wrote the above test on a scale of 1 to 4 on the basis of the extent to which they demonstrate the behaviors used in the definition of the attributes (see the instructions to the teachers in Appendix E). The high student group consists of the top twenty percent of the students in each class.

Low Student Group: The low student group consists of the bottom twenty percent of the students from each class (on the basis of the teacher ratings noted above).

Test Of Likert Items (TOLI): This test consists of twenty-five opinion statements relevant to the affective attributes which are assessed in the present study (see Appendix C). Students are asked to respond to these items on a Likert scale with the following response categories: strongly agree, partly agree, partly disagree and strongly disagree.

V. Questions and Hypotheses

A rationale for the construction of the test items in this study will be developed in Chapter III. Whether or not this rationale is a feasible approach to evaluation in the affective domain of science education is the major question which this study has been designed to answer. The validity and stability of the test items constructed on the basis of this rationale will be examined. This question will also be discussed in terms of the problems encountered when applying the rationale to test-item construction.

Test-Stability: Test-retest correlation coefficients will be obtained for the Test On Scientific Attitude and its two subtests (Appendix B).

Construct Validity: The validity of the test items will be examined under the three components of construct validity (substantive, structural, and external) defined by Loevinger (1967, pp. 92-108). These three components incorporate the concepts of content, construct, predictive and concurrent validity discussed by other writers (Magnusson, 1966, pp. 127-137; Cronbach and Meehl, 1955). The concept of test homogeneity is also included in the structural component.

1. Substantive Component: The content validity of the test items will be argued on the grounds that a panel of judges will be used to define the attributes in terms of student behaviors. These behaviors will be within the context of science activities. A panel of judges will also be used in selecting the keyed response for each test item.

2. Structural Component: Item analysis will be used

to provide an estimate of the homogeneity of the test items. The empirical structure underlying the test items, as indicated by a factor analysis solution, will be compared with the structure predicted from the behavioral definations of the attributes.

3. External Component: The students were divided into a high and low group (see the definition of terms) using teacher ratings of classroom behavior. The rejection of the following null hypothesis will lend support to the claim of concurrent validity.

Hypothesis I: There is no significant difference between the mean score of the low student group and the mean score of the high student group on TOSA, CCS, ICS and TOLI when scholastic ability as measured by the Cooperative School and College Ability Test (SCAT) is the covariate.

Correlations of test scores with teacher ratings will also be reported.

Descriptive Statistics: A number of descriptive statistics such as correlations, means, and standard deviations will be reported for the Test On Scientific Attitude and the Test Of Likert Items. The following null hypotheses will also be tested:

Hypothesis II: There is no significant difference between the mean score of males and the mean score of females on TOSA, CCS, and ICS when the covariates are scholastic ability and reading ability.

Hypothesis III: There is no significant difference in the TOSA, SCAT, and STEP reading scores between students writing in the spring and students writing in the fall.

VI. Delimitations

The present study examines only one approach to the evaluation of affective objectives (the rationale outlined in Chapter III). This rationale will be discussed in terms of its usefulness in the development of evaluation instruments. No attempt will be made to make a comparative evaluation of the rationale developed in this study against some other approach.

The present study is confined to eight of the sixty-five attributes listed in the Nay-Crocker inventory. These eight are critical mindedness, suspended judgement, respect for evidence, honesty, objectivity, willingness to change opinions, open-mindedness, and questioning attitude. These eight were chosen because they are grouped together in the inventory and because the development of these characteristics in students is generally accepted as an important objective of science education (Alberta Department of Education, 1970). Because this is primarily a methodological study no attempt was made to deal with the complete inventory.

The present study does not investigate the problem of which teaching methods and materials are most appropriate to foster attitude development and change. Although the importance of research into this area is recognized, these topics are beyond the scope of this study.

VII. Limitations

Because of the limited sample (only grade eleven chemistry and physics students will be tested), the present study has limited value with respect to generalizabiltiy of the results. But since this is basically a methodological study, the main question of interest is whether or not the constructed test items accurately identify selected characteristics of the students in the sample. The test scores obtained in this study are not used to make statements concerning the characteristics of science students in general. Grade eleven science students were selected for the purpose of this study because it was felt that the attitudes of these students would be fairly stable.

VIII. Experimental Procedures and Design

A list of behavioral objectives was compiled for each attribute and a panel consisting of professors and graduate students in science education in the Department of Secondary Education at The University of Alberta rated each behavior in terms of its relevance to a particular attribute (see the instructions to the panel in Appendix D). The list of behaviors defining each attribute was reduced on the basis of the panel's responses. These behavioral definitions were used as guidelines for the construction of test items.

TOSA (Appendix B) was administered to grade eleven chemistry and physics students during the spring semester. These students also wrote the Watson-Glaser Critical Thinking Appraisal which is described in Chapter III and TOLI (Appendix C). During the fall semester, TOSA was administered twice, over a three-week period, to a second group of grade eleven chemistry and physics students. Teacher ratings (Appendix E) were obtained only for those students who wrote the test during the spring semester.

Hypothesis I was tested by a one-way analysis of covariance in which the factor levels are the high and low student groups based on the teacher ratings and the covariate is general scholastic ability. Only those students who wrote during the spring semester were included in this analysis. One-way analysis of covariance in which scholastic ability and reading ability are the covariates was used to test Hypothesis II. Hypothesis III was tested by the use of independent-sample t tests.

Factor Analysis was used to examine the underlying structure of the test items, and item analysis was used to study the properties of the individual items. A more detailed discussion of these experiemental and analytical procedures appears in Chapter III.

CHAPTER II

Review Of The Related Literature

The literature review presented in this chapter is discussed under the following topics: definitions of the affective domain, critical discussion of behavioral objectives, techniques used in attitude measurement, and specific science-attitude scales which have been developed.

I. Defining the Affective Domain

Introduction: The set of objectives which are generally classified within the affective domain involve the development of interests, attitudes, values, appreciations, and adjustments (Bloom, et. al., 1956, p. 7). Various approaches to categorizing, summarizing and defining this set of objectives are discussed below.

A Taxonomy of Affective Objectives: The taxonomy of affective objectives in education which was developed by Krathwohl, Bloom, and Masia (1964) is a general classification scheme, that is, it is not defined in terms of any one subject area. This taxonomy defines the affective domain in terms of a valuing system. The term valuing refers to the tendency to recognize certain objects or activities as being worthy of an individual's attention. The taxonomy has five main categories and each of these is subdivided further. These categories describe levels of internalization of values proceeding from simple awareness of stimuli, to the formation of values concerning these stimuli, to the inclusion of these values in an overall philosophy of life.

The authors provide a number of examples of objectives under each category and they also provide examples of test items to determine whether or not these objectives have been met. However, most of these examples deal with art, music, and literature appreciation. There are only occasional examples from science. A possible reason for this preponderance of objectives relevant to the arts may be that these objectives more readily fit into the classification scheme of the taxonomy than do the affective objectives in science. This may also be an indication that this scheme is not readily applicable to the definition of affective objectives for science education.

Eiss (1968), in his report on the NSTA conferences on scientific literacy, describes an attempt to make use of the categories in this taxonomy in summarizing the affective characteristics which define a scientifically literate person. These characteristics are listed under the categories of awareness of conditions, acceptance of values, and preference for values. However, the progressive internalization process described by Krathwohl and his co-authors is not demonstrated in these lists. The distinction between the three categoties is not clear. For example, the following two statements are quite similar in meaning:

"recognizes that the achievements of science and technology properly used are basic to the advancement of human welfare" and "realizes that science is a basic part of modern living". The first one is listed under awareness of conditions and the second one is listed under acceptance of values. Characteristics listed in one category do not always have a corresponding characteristic at a different level of internalization in the other two categories. The participants at the conference were not able to identify any objectives at the higher levels of the taxonomy of affective objectives.

The above attempt to apply the taxonomy to science objectives is a further indication that the affective objectives in science education do not readily lend themselves to the classification scheme outlined in the taxonomy.

In the initial discussion of the three domains of educational objectives (cognitive, affective, and psychomotor), the affective domain was defined in terms of interests, adjustments, attitudes, appreciations, and values (Bloom, et. al., 1956, p. 7). These categories were not used in the later development of the taxonomy of affective objectives because the authors of this taxonomy felt that the variety of meanings associated with these terms rendered them inadequate to serve as a basis for the construction of a continuum (Krathwohl, et. al., 1964, p. 24). However, Nay and Crocker (1970) claim that these terms can be defined quite specifically if they are considered in the context of science work. Their approach to the definition of the affective domain in science education is presented in the next section.

The Affective Domain Defined in Science: Nay and Crocker (1970, pp. 61-62) have developed an inventory of affective attributes of scientists. This is a list of interests, attitudes, adjustments, appreciations, and values which scientists are generally expected to demonstrate in their work. This list was compiled through an extensive review of the literature on the nature and philosophy of science and from information obtained through interviews with scientists.

They feel that these attributes are "primarily dictated by the nature of scientific inquiry and are operationally definable for scientists", and that these attributes are fundamental to a person's decision to become a scientist and to his work as a scientist. They contend that these attributes are essential to the pursuit of science. Therefore, students should be led to understand the significance of these attributes to the scientist in his work and should also be encouraged to demonstrate these attributes in their own activities. If these attributes are defined in terms of student behaviors, a list of behavioral objectives for science education can be developed from this inventory.

A complete distinction between attitudes and adjustments is not made in this inventory. Nay and Crocker have identified a set of operational adjustments, "behaviors which underlie competence and success in science", and a set of intellectual adjustments, "behaviors which are foundational to the scientist's contribution to or acceptance of new scientific knowledge". The term attitude is associated with the list of intellectual adjustments. The attributes which will be examined in the present study are listed in this category.

The application of this inventory to the present study is discussed in Chapters I and III.

Defining Attitude: The definition of attitude as a predisposition to respond in some preferential manner with respect to some object or situation has been widely accepted for several years (Allport, 1935, p. 8; Fishbein, 1967, p. 477; Rokeach, 1968, p. 112). It is also generally accepted that attitudes are learned from experiences involving the attitude object or situation. However, there is some disagreement concerning the various components of which attitudes are comprised. Rokeach defines attitudes in terms of three components (cognitive, affective, and behavioral). These three components represent knowledge about the attitude object, a tendency to take a positive or negative position toward the attitude object, and some type of observable action with respect to the attitude object.

Fishbein prefers to consider attitude as a unidimensional concept. In his definition, attitude represents only the tendency to take a positive or negative position toward the attitude object. He defines beliefs as knowledge about the attitude object, and behavior as the overt action stimulated by encounters with the attitude object.

Nay and Crocker define two components of the affective attributes listed in their inventory. These correspond to the cognitive and behavioral components defined by Rokeach. The cognitive component represents the student's understanding of the role of the scientific attributes in the activities of scientists. The behavioral component represents the tendency for the student to demonstrate these attributes in his own science work.

In the present study, a multidimensional approach incorporating the concepts presented by Nay & Crocker and Rokeach will be used to define three components of the attributes in the Nay-Crocker inventory. Rokeach's terms, "affective component" and "behavioral component", are somewhat ambiguous when used in the context of this study. The term, affective, has already been used to refer to a set of attributes of scientists which have been extrapolated to the affective domain in science education. Since these attributes will be defined in behavioral terms, this is a possible source of confusion with Rokeach's behavioral component.

To avoid confusion which might arise from the dual use of these terms, Rokeach's three components will be referred to as the cognitive, intent, and action components in this study. The cognitive component represents the student's understanding of the significance of the attribute to the scientist in his work. The intent component represents the student's tendency to show approval or disapproval of behaviors which define the attribute. This will be indicated by the student's endorsement of specific responses in situations relevant to the attribute. The action component refers to the extent to which the student actually demonstrates the behaviors which define the attribute if placed in a position to do so.

II. Behavioral Objectives

Behavioral objectives are educational objectives which describe observable behaviors that students are expected to demonstrate as a result of participation in a planned activity in the classroom (McAsham, 1970, p. 8). Objectives stated in terms of student behaviors are also referred to as instructional objectives (Eisner, 1969; Mager, 1962; Popham, 1969). Mager (1962, p. 12) insists that behavioral objectives also specify the conditions under which the behaviors should be initiated and the desired minimal level of learner performance. Popham (1969, p. 35) agrees that these may be useful additional considerations, but he stresses that the most important criterion which must be met is the specification of observable student behaviors.

Popham (1969, p. 37) claims that general objectives which do not specify student behaviors are of "almost no use to the teacher". Plowman (1971, p. xxvii), on the other hand, feels that behavioral objectives are not inherently better than non-behavioral objectives and that all types of objectives (general and specific; behavioral and non-behavioral) contribute to the overall planning of educational activities. General objectives can be very useful in guiding long-term planning to provide a common theme in a teacher's approach to teaching a particular subject or unit. However, Plowman recognizes that objectives must be translated into observable and measureable functions before they can serve a useful diagnostic, prescriptive, and evaluative purpose in the direction and assessment of learning.

Eisner (1969, pp. 14-15) makes a distinction between instructional (behavioral) objectives and expressive objectives. He discusses behavioral objectives in terms of their application to curriculum development and revision. Desired student behaviors are defined. Materials and activities which are predicted to be useful in developing these behaviors are then selected. Revisions are made on the basis of the results of evaluation designed to determine the extent to which the behaviors are achieved by students.

Expressive objectives describe an educational situation which includes a problem, and a number of situations and tasks related to the problem. The objective is the outcome which results from student participation in these situations, but the expected outcome behavior may not be the same for all students. For example, not all students would be expected to give the same interpretation to a piece of literature. Eisner feels that expressive objectives are particularly applicable to the arts. He expresses a fear that the use of lists of prescribed outcome behaviors may cause the teacher to miss opportunities to pursue open-ended situations arising in the classroom and to neglect the individual differences of his students.

In his discussion of the literature on behavioral objectives, McAsham(1970, p. 6) states that the main criticism against the use of behavioral objectives is that some teachers may become alienated because of the degree of specificity required in the writing process. Haberman (1968, p. 93) claims that excessive dependence on behavioral objectives may cause those objectives and subject areas which are most easily specified in behavioral terms to be given undeserved prominence. In particular, the formation of generalizations may be neglected and the development of skills may be overemphasized.

Most of the criticisms of the use of behavioral objectives take the form of objections against sole use of behavioral objectives and total disregard for non-behavioral objectives at all levels of educational planning. The fact that objectives are stated in behavioral terms does not necessarily mean that they are better or more important than non-behavioral objectives. In his summary of the objections against behavioral objectives, McAsham (1970, p. 7) points out that those individuals who criticize certain features of behavioral objectives also admit some of their advantages in research, curriculum development, and classroom instruction. However, behavioral objectives must always be written at an appropriate level of specificity so as to avoid unrealistic and impractical objectives. Lists of behavioral objectives should be screened and appropriately grouped so that long lists of trivial behaviors are not included.

If specific behavioral objectives are associated with broader, general objectives, they may become more meaningful and may gain greater acceptance by a greater number of people. This approach will be used in the present study. Each objective will be stated at two levels. For example, a general objective is to encourage students to demonstrate suspended judgement in science work. Suspended judgement will then be defined in terms of student behaviors. This objective can be stated in the following way: to develop the attitude of suspended judgement by encouraging the student to generalize only to the degree justified by available evidendence, to recognize conclusions as being tentative . . . The complete list of behavioral objectives defining suspended judgement is given in Chapter III.

When appropriate behavioral objectives were found in the literature, these were included in the initial list of behaviors (Appendix D) submitted to the panel for evaluation (Diederich, 1967; Eiss and Harbeck, 1969; Obourn and Johnson, 1960).

III. Techniques Used in Attitude Measurement

Thurstone Scales: Much of the early work in attitude measurement was done by Thurstone (1928; 1931). He defines opinions as the verbal expressions of an attitude. An individual's attitude toward some object or stiuation is inferred from his opinions directed to that object or situation.

The attitude scales which Thurstone constructed consist of a list of opinion statements directed toward some specific attitude object. Each statement is assigned a number. An opinion statement with a high number represents a strong positive position with respect to the attitude object. A statement with a low number indicates a strong negative position. The most positive and most negative statements from Thurstone's scale on attitude toward negroes are given below with the corresponding scale values (Shaw, 1967, p. 363):

0.9 - The negro will always remain as he is - a little higher than animals.
10.3 - I believe that the negro deserves the same social privileges as the white man.

The respondent is asked to mark those statements with which he agrees and his position with respect to the attitude object is indicated by the average value of the numbers assigned to the statements that he has marked. Data collection and analysis procedures used to assign scale values to the opinion statements are discussed by Thurstone (1928, pp. 82-88) and by Torgerson (1958, pp. 159-246).

Thurstone's scaling procedures have been widely accepted and extensively used in attitude measurement. Thurstone has developed a sound theoretical and mathematical foundation to support the analytical procedures which are used in the calculation of the scale values.

In the construction of a scale by the procedures outlined by Thurstone, a unidimensional attitude ~~object~~ is assumed. Therefore, a large number of scales would be required to identify all of the dimensions of the affective domain in science education. Since a considerable amount of work is required on the part of the respondents who provide the data from which the scale is to be determined, the construction of a large number of scales may not be a practical undertaking.

Likert Scales: The attitude-measurement technique developed by Likert (1932) has been widely used and a large portion of the attitude scales which have been constructed are of this type. This technique also makes use of a list of opinion statements regarding the attitude object. Respondents are asked to check one of the following categories for each statement: strongly agree, agree, uncertain, disagree, and strongly disagree. Each response is scored on a scale of 1 to 5 where strong agreement with a positive statement and strong disagreement with a negative statement are scored 5. Other response categories such as approve-disapprove, like-dislike, etc. are also used.

A number of weaknesses of the attitude instruments employing Likert scales have been identified. The response biases associated with Likert scales are discussed later in this chapter. Responses of Likert items may also be affected by differing meanings which different respondents may identify with the response categories. For example, different respondents may assign different meanings to terms such as partly agree, strongly agree, sometimes, often, etc. Although the response categories are scored by integer values from 1 to 5, no measures are taken to ensure that the distances between these categories are consistent across the scale. For example, the distance between strongly agree and agree may not be the same as the distance between agree and uncertain. These distances may vary from one respondent to the next.

Semantic Differential: This technique requires the respondent to rate a concept on a continuum of opposites. The positions along the continuum are assigned numbers which usually range from 1 to 5 or 1 to 7. For example, science is good.....bad. In this case, the space closest to good would be assigned a value of 5. If the word science is rated on a large number of such scales, a total rating can be obtained.

The semantic differential was initially designed for the measurement of meaning (Osgood, *et*. *al*., 1957). Through factor analysis, it was possible to identify a group of scales which was strongly evaluative in nature. Some of the scales included in this group are good-bad, fair-unfair, and valuable-worthless. This set of scales has been used to obtain a measure of attitude toward the church, negroes, and capital punishment, and it has been suggested that this set of scales can be used to obtain a measure of attitude towards any specific object (Osgood, *et*. *al*., 1957, pp. 189-216).

This approach is not appropriate for the present study, because the information obtained is not directly related to the classroom situation. Therefore, this information may have only limited usefulness in the process of curriculum development and evaluation.

Other Methods: In addition to the three techniques already mentioned, Oppenheim (1966, pp. 120-154) and Shaw (1967, pp. 21-32) discuss a number of less commonly used methods of attitude measurement. These include questionnaires, interview schedules, and various indirect techniques such as sentence-completion, picture-interpretation, and word-association. Error-choice questions have also been used in attitude measurement. Here the respondent must choose between two equally erroneous alternatives to a question. One alternative errs in a favorable direction with respect to the attitude object while the other alternative errs in a negative direction.

Response Biases Associated With These Techniques: Various test formats are subject to certain response biases. That is, some individuals have marked tendencies to give a certain type of response regardless of the content of the question (Cronbach, 1946; Cronbach, 1950). In these articles, Cronbach cites a number of research studies to support his arguments. Acquiescience refers to the tendency to respond with like rather than dislike, agree rather than disagree, true rather than false, etc. Some individuals show a greater tendency to go to extremes. This tendency will affect responses on the semantic differential and on Likert scales. The Likert scale is also subject to the tendency to remain uncommitted resulting in some individuals giving a large number of "uncertain" or "undecided" responses. This response bias may be avoided by deleting the uncertain or undecided category from the scale. Open-ended techniques such as sentence-completion, and picture-interpretion may be subject to the bias of inclusiveness.

Some individuals tend to write down everything that they know and feel while others write down only a selected quantity.

Discussion: The position taken in this study is that, since attitudes are defined as predispositions to some preferred response with respect to the attitude object, an individual's attitude can be inferred from his endorsement of certain courses of action in situations relevant to the attitude object. A test-item format which is appropriate for this approach is a multiple-choice item in which the stem describes a situation relevant to the attitude object and the distractors describe alternate courses of action. The test items in TOSA (Appendix B) are of this format. TOLI (Appendix C) has been included in this study to provide a comparisson of Likert-scale items with test items of the above format.

IV. Summary of Research on Attitude Measurement In Science Education

Test On Understanding Science: The intention of this test which was developed by Klopfer and Cooley (1961) is to measure understandings about the nature of the scientific enterprise, scientists, and the methods and aims of science. A list of themes are described to provide definitions for these three dimensions. The test items in this test are four-alternative, multiple-choice items.

A panel of consultants was used to establish the content validity of the test items and of the themes. The validity of the test items was further demonstrated in a study involving a group of students who were in active contact with working scientists over a two-month period. This group of students plus a control group, who did not interact with scientists, were tested at the beginning and end of this two-month period. The experimental group showed a significant gain in their test scores while the control group did not. The KR-20 reliability coefficient reported for form X of this test is 0.76 for a sample of 2535 high school students who wrote the test during the fall of 1960.

This test has been extensively used in research studies which have attempted to identify factors which might foster the development of student understandings of science, scientists, and science processes. *The Seventh Mental Measurements Yearbook* (Buros, 1972, p. 804) cites thirty-three studies which make use of TOUS.

A Test To Measure "The Scientific Attitude": Noll (1935; 1936) developed a test to measure the following characteristics which he defined as identifying "the scientific attitude": accuracy in operations, intellectual honesty, open-mindedness, suspended judgement, looking for true cause and effect relationships, and criticalness. The questions in this test are mainly true and false questions. Following are some examples from the test:

> Evolution is something I don't care to know about.
> People with red hair are usually ill-tempered.
> If one of my teachers says a thing is so, it must be so.

Other questions require the student to record observations from diagrams. Some multiple-choice questions are also included. Noll reports the split-half reliabiltiy coefficient of 0.80 for the 135-item test based on a sample of 383 students from grades eight to twelve.

Kahn (1962) used this test in his study on the use of current events in science to develop scientific attitude. The author of the present study was not able to find any other studies in which this test was used.

Projective Test Of Attitudes: Lowery (1966) makes use of indirect techniques to measure student attitude toward science, scientists and science processes. He does not provide detailed definitions of the above. The test consists of three subtests. The first subtest is a word-association test. The second subtest is a picture-interpretation test in which students are shown a picture and are asked to describe what lead up to the scene, what is happening in the scene, what the feelings of the characters are, and what the outcome will be. The following are examples of the type of pictures used in this test; a student meeting a scientist, a student reading a science headline, and a student looking at some laboratory equipment. The third subtest is a sentence-completion test. For example, "The field of science is ...." All three subtests contain questions in all three areas. The test is subjectively scored by rating each response as positive, neutral or negative. The author of the present study was not able to locate any examples of the use of this test in research.

An Inventory Of Scientific Attitudes: Moore and Sutman (1970) define three intellectual attitudes (based on some knowledge of the attitude object) and three emotional attitudes (based on feelings or emotional reactions) toward science. Each attitude is stated both positively and negatively. An example of a positive, intellectual attitude is "The laws and/or theories of science are approximations of truth and are subject to change." An example of a positive emotional attitude is "Science is an idea generating activity. It is devoted to providing explanations of natural phenomena. Its value lies in its theoretical aspects."

The test consists of sixty opinion statements related to the attitudes referred to in the above paragraph. Students are required to respond to each statement on a Likert scale consisting of the following four response categories: agree strongly, agree mildly, disagree mildly, disagree strongly. The validity of the test was demonstrated in a study involving a control group which received regular classroom instruction and experimental groups which received instruction directed toward the development of the attitudes measured by the test. The means of the pre test and post test scores were tested for each group by the use of correlated-t tests. The control group showed a significant drop from pre test to post test while the experimental groups showed significant gains. A test-retest correlation coefficient of 0.93 was obtained for the twenty-three students in the control group. Following are examples of statements from this test:

> There is no need for the public to understand
> science for scientific progress to occur.
> A major purpose of science is to produce new drugs
> and save lives.
> One of the most important jobs of a scientist is
> to report exactly what his senses tell him.
> Scientists do not have enough time for their
> families or for fun.

Lauridsen and LaSheir used a revised version of this test in their study of the effect of ISCS on affective characteristics of students. Other examples of the use of this test in research were not found.

A Science Support Scale: This scale which was developed by Schwirian (1968) is based on Barber's (1962) summary of five cultural values which he considers to be conducive to the development of positive scientific attitudes. These five values are rationality (acting on the basis of available evidence), utilitarianism (interest in natural phenomena), universalism (judging scientists only on the basis of their qualifications), individualism (commitment to individual conscience), and meliorism (acceptance of the benefits of science). The scale consists of forty opinion statements designed to measure an individual's support of these values. The five Likert-response categories range from strongly agree to strongly disagree. A neutral category is included. A split-half reliability coefficient of 0.87 is reported for a sample of 196 non-science majors at univerisity. Claims for item validity are made on the basis of consistent item to total score relationships. Following are examples of opinion statements from this scale:

> The skepticism of the scientist should be limited to his work.
> In the long run, man's lot will be improved by scientific knowledge.
> Those who have a history of mental illness cannot be trusted to do important scientific work.
> There is no place in science for sexual deviants such as homosexuals.
> The questions which are really important to man cannot be answered by science.

The author of the present study was not able to locate any examples of the use of this test in reaseach.

Attitudes Toward Science and Scientific Careers: Allen (1959) developed a scale to measure attitudes toward science and scientific careers. This scale consists of ninety-three opinion statements which pertain to characteristics of scientists, the nature of science work and the contributions of science to mankind. The Likert response categories used in this scale are completely agree, partial agreement, neutral, partially disagree, and totally disagree. Following are examples of statements from this scale:

> Science is not sufficiently appreciated by most people.
> Science is a systematic way of thinking.
> Scientists are seldom concerned with their working conditions.
> Scientists have unusually intelligent mothers.
> Friends often discourage girls from taking high school science courses.

The author of the present study was not able to locate any examples of the use of this test in research.

Discussion: Most of the tests described above have fairly high reliability coefficients and reasonable attempts to demonstrate test-validity have been made in most cases. However, some of these tests make use of Likert scales and are subject to the response biases associated with this type of scale. Cronbach (1950, p. 4) claims that the effect of response biases may result in spuriously high reliability coefficients. That is, the test items may be consistently measuring the response bias rather than the dimensions which they were designed to measure.

The major criticism of these attitude scales that the present study has to make is that the definitions on which these scales are based are usually too general to serve a useful purpose for curriculum development. This often results in an attempt to include a wide variety of dimensions in one test (interest, processes, values, attitudes, and knowledge about the characteristics of scientists).

Nay and Crocker (1970, p. 65) criticize the above science attitude scales because behavioral objectives were not used in defining the dimensions of the scales, because the scales do not discriminate between the affective and cognitive component of the attitudes being measured, and because the content often does not adequately represent classroom situations and experiences.

Chapter III

Experimental Procedures And Design

I. A Rationale for the Construction of the Test Items

Introduction: The following points were discussed briefly in Chapter I under the background of the problem: purpose for evaluation of the achievement of affective objectives, choice of a set of objectives to define the affective domain in science education, definition of the objectives, and selection of an appropriate test format. A survey of the literature related to these topics was presented in Chapter II. The following elaboration of the points made in Chapter I, with reference to the relevant ideas in Chapter II, provides the rationale for the construction of the tests items in the present study.

Purpose of Evaluation: The approach taken to test-item construction in the present study was influenced by the position of evaluation in the overall model for educational planning illustrated in Figure I on page 36 (Engman, 1968, p. 87). The model incorporates evaluation as a check on the effectiveness of the methods and materials used at phase II in achieving the objectives defined at phase I. The evaluation provides information to guide the analysis and revision represented at phase IV. If this information is to serve a useful evaluative purpose, the objectives at phase I must be stated in terms of observable student behaviors.

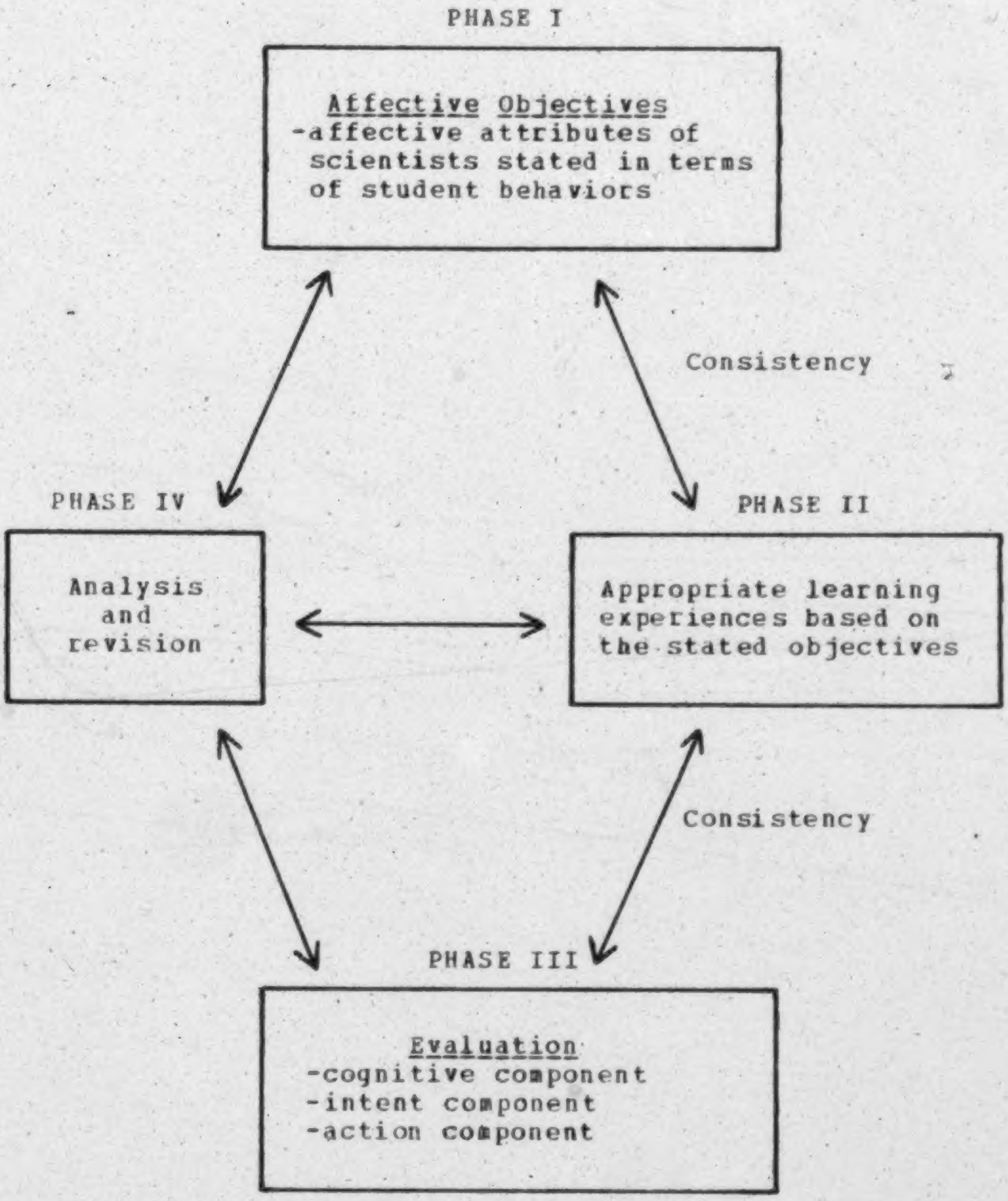
PHASE I
Affective Objectives
-affective attributes of scientists stated in terms of student behaviors
Consistency
PHASE IV
Analysis and revision
PHASE II
Appropriate learning experiences based on the stated objectives
Consistency
PHASE III
Evaluation
-cognitive component
-intent component
-action component

FIGURE I. A Model for Educational Planning

This point is generally supported by the literature on behavioral objectives presented in Chapter II. As is indicated later in this chapter, the first step in the present research is to obtain behavioral specifications of the affective attributes.

The present study does not investigate phases II AND IV of the model. Regardless of what objectives a teacher has in mind and the activities used in the science classroom to achieve them, students will develop certain attitudes. Therefore, it is possible to examine methods of identifying those attitudes which a given group of students possess without investigating the dynamics of attitude development and change.

Selection of Affective Objectives: The Nay-Crocker inventory of affective attributes of scientists (Appendix A) is used in this study as a summary of general affective objectives for science education. Following are examples of how these general objectives can be stated: to develop an understanding of the relationship between science and technology, to develop objectivity in science work, to develop a desire for understanding of natural phenomena, and to develop an appreciation for the strengths and limitations of science. The present study is confined to eight of the attributes listed under the heading of attitudes or intellectual adjustments. These are objectivity, open-mindedness, honesty, suspended judgement (restraint), respect for evidence (reliance on fact), willingness to change opinions, critical mindedness, and questioning attitude. Since the present study is basically a methodological one, no attempt was made to deal with the complete set of attributes. These eight were selected because they are generally accepted as desirable objectives for science education (Alberta Department of Education, 1970) and because they are grouped together in the inventory under attitudes. There is a larger foundation of research on attitude measurement than on the measurement of values, appreciations, etc. The process by which the objectives have been defined in more specific terms is outlined in the next section.

Definition of the Objectives: Definition of the objectives in behavioral terms is consistent with the main purpose of evaluation expressed above and is also consistent with the approach to attitude measurement taken in this study (discussed below under test-format). In his discussion of behavioral objectives, McAsham (1970, p. 4) indicates that "The primary reasons for the current emphasis upon writing behavioral objectives are to: (1) aid in curriculum planning, (2) promote increased pupil achievement, and (3) improve the techniques and skills of program evaluation".

A list of behaviors defining each of the attributes was compiled. A panel consisting of 3 faculty members and 8 graduate students in the Department of Secondary Education at The University of Alberta rated each behavior on the scale of 0-1-2 in terms of its importance in defining a specific attribute. The instructions to the panel members and the original list of behaviors are given in Appendix D. The distribution of the panel responses and a total rating for each behavior are included in this appendix.

The decision to retain a behavior in the final definition was made on the basis of the total rating given to that behavior and the distribution of the ratings for that behavior. After a general inspection of the distributions and totals, the decision was made to retain all behaviors with a total of fourteen or larger. Behaviors which received a total rating of thirteen were also included in the final definitions if the distribution of ratings showed a consensus among the panel members. For example, a behavior with a distribution of 0 9 2 was included while a behavior with a distribution of 3 3 5 was not included.

The list of behaviors retained to define these attributes are listed below with the distribution of ratings for each behavior. The first column contains the number of panel members who rated the behavior to be trivial or not related to the attribute under which it is listed. The second column contains the number of panel members who rated the behavior to be an important defining characteristic of the attribute. The third column contains the number of panel members who rated the behavior to be a very important defining characteristic of the attribute. The fourth column is the total rating (0 times column 1 plus 1 times column 2 plus 2 times column 3):

A student demonstrates critical mindedness when he:

0	1	10	21	-looks for inconsistencies in statements and conclusions
1	6	4	14	-consults a number of authorities when seeking information
0	3	8	19	-looks for empirical evidence to support or contradict explanations
1	6	4	14	-asks many questions starting what, where, why, when and how
1	2	8	18	-challenges the validity of unsupported statements

A student demonstrates suspended judgement (restraint) when he:

1	4	6	16	-generalizes only to the degree justified by available evidence
1	3	8	19	-collects as much data as possible before drawing conclusions
1	3	7	17	-recognizes conclusions as being tentative
0	9	2	13	-consults several authorities (texts, periodicals, people) before drawing conclusions

A student demonstrates respect for evidence (reliance on fact) when he:

0	2	9	20	-looks for empirical evidence to support or contradict explanations
1	7	4	15	-collects as much data as possible before drawing conclusions
0	6	5	16	-demands that explanations fit the facts
0	2	9	20	-demands supportive evidence for unsubstantiated statements
0	5	6	17	-supplies empirical evidence to support his statements

A student demonstrates honesty when he:

0	2	9	20	-reports observations even when they contradict his hypotheses
0	6	5	16	-acknowledges work done by others
1	4	6	16	-considers all available information when

forming generalizations and drawing conclusions

A student demonstrates objectivity when he:

0	0	11	22	-considers all available data (not only that portion which supports his prior hypotheses)
0	3	8	19	-reports observations even when they contradict his hypotheses
0	7	4	15	-considers and evaluates ideas presented by others
1	5	5	15	-examines many sides of a problem and considers several possible solutions
0	7	4	15	-considers both pros and cons when evaluating a situation

A student demonstrates willingness to change opinions when he:

1	3	8	19	-recognizes conclusions as being tentative
0	8	3	14	-recognizes that knowledge is incomplete
0	7	4	15	-considers and evaluates ideas presented by others
0	8	3	14	-evaluates evidence which contradicts his hypotheses
0	2	9	20	-alters his hypotheses when necessary to accommodate empirical data

A student demonstrates open-mindedness when he:

0	7	4	15	-considers and evaluates ideas presented by others
0	7	4	15	-evaluates evidence which contradicts his hypotheses
1	5	5	15	-considers several possible options when investigating a problem
0	5	6	17	-considers both pros and cons when evaluating a situation

A student demonstrates questioning attitude when he:

0	4	7	18	-looks for inconsistencies in statements and conclusions
0	9	2	13	-consults a number of authorities when seeking information
0	5	6	17	-looks for empirical evidence to support or contradict explanations
2	2	7	16	-asks many questions starting who, what, where, why, when and how
1	2	8	18	-challenges the validity of unsupported statements

The behaviors which were retained to define critical mindedness are the same as those retained to define questioning attitude. For the remainder of this study critical mindedness will be used to refer to this set of behaviors. This situation also applies to objectivity and open-mindedness. The term objectivity will be used to refer to the set of behaviors defining these two attributes. Disciplined thinking, which is included in Appendix D, is not included in the above list because test questions to measure this attribute were not constructed. The behaviors which were indicated to be defining characteristics of disciplined thinking appear to be process oriented. For example, organization of data and distinguising between relevant and non-relevant data.

Description of the Test Format Used in the Present Study: Nay and Crocker (1970, p. 65) criticize existing science-attitude scales because behavioral objectives were not used in defining the dimensions of the scales, because the scales do not discriminate between the affective and cognitive components of the attitudes being measured, and because the content of the scales often does not adequately represent classroom situations and experiences. Attempts have been made to deal with the above points in the present study.

The behavioral specification of the affective objectives in this study is consistent with the definition of attitude presented in Chapter I. Attitudes were defined as predispositions to some preferred response. The approach to attitude measurement taken in the present study is that an individual's attitude can be inferred from his or her endorsement of various courses of action in certain situations relevant to the attitude being measured. The test-item format used in the present study is a multiple choice question in which the stem presents a situation and the distractors are four different courses of action pertaining to the situation. One of the courses of action (the keyed response) is consistent with one of the behavioral specifications used in defining the attributes.

In the following question (question 2 from Appendix A), the stem describes a situation from the point of view a scientist and the four alternatives describe four courses of action that the scientist could take:

> A science magazine reports that a scientist produced a type of water that boils at 450°F under one atmosphere of pressure. Another scientist reading this report would probably
>
> A. believe the report if it was written by a highly respected scientist.
> B. disbelieve the report because he would know that water boils at 212°F under one atmosphere of pressure.
> C. do experiments to try to prove that it was wrong.
> D. neither believe not disbelieve the report until other scientists study this problem.

The keyed response for this question is D. This alternative is consistent with the behavior, consults several authorities (texts, periodicals, people) before drawing conclusions.

The forty items which were constructed as a part of the present study are divided into two subtests of twenty items each (Appendix B). The stems of the items in the Cognitive Component Subtest describe a situation which a scientist might encounter in his work. The student is asked to select the course of action which is most appropriate for the scientist. The question given above is an example of a test item from this subtest. This item is designed to measure the student's understanding of the role of suspended judgement in influencing a scientist's actions.

The stems of the items in the Intent Component Subtest present a situation which the student may encounter in the science classroom or in every-day activities. The student is asked to select a course of action which best describes his reaction to this situation. The following question (question 24 from Appendix B) is an example of an item from this subtest:

> "Light travels as a stream of particles."
> "Light travels as a wave."
> If you came across these two statements in two different science books, which of the following would you do?
>
> A. Ask your teacher to tell you which statement to accept.
> B. Check other science books for statements on this topic.
> C. Assume that scientists are not certain as to how light travels.
> D. Accept the statement in the newer book.

The keyed response for this item is B. This alternative is consistent with the behavior, consults a number of authorities when seeking information, which is listed under critical mindedness and suspended judgement. The twenty questions constructed for each subtest are reported in Appendix B. The following list gives a summary of those questions which were designed to measure each of the attributes. Items 1 to 19 are from the cognitive subtest and items 21 to 40 are from the intent subtest:

Critical mindedness (questioning attitude) -
9, 19, 21, 24, 25, 31, 32, 36

Suspended judgement (restraint) -
1, 2, 4, 6, 7, 11, 18, 25, 26, 27, 28, 34, 35, 37

Respect for evidence (reliance on fact) -
10, 11, 12, 14, 16, 26, 27, 28, 31, 32, 33, 38

Honesty -
3, 8, 18, 22, 33, 39

Objectivity (open-mindedness) -
8, 13, 15, 17, 23, 29, 30, 36, 40

Willingness to change opinions -
1, 3, 4, 5, 6, 10, 29, 30, 37

When the items were writen, an attempt was made to distribute the items evenly among the six attributes and between the two subtests (CCS and ICS). However, since serveral of the behavioral objectives defining the attributes are listed under more than one attribute, the questions based on these behaviors are listed under more than one attribute. Another factor contributing to the uneven distribution of questions among the attributes is that the items were associated with only one behavior when they were written. Closer inspection of the items revealed that, for some questions, the keyed response was related to one attribute while other alternatives were related to other attributes.

A large amount of science testing material was surveyed in an attempt to find test questions of the type described above. This search was not very productive. Some of the questions in the Test On Understanding Science (Klopfer and Cooley, 1961) were found to be relevant to the Cognitive Component Subtest, but examples of test items relevant to the Intent Component Subtest were not found. Some of the ideas in questions in TOUS were used in the construction of some of the questions in the Cognitive Component Subtest. Although appropriate test questions were not found in the science literature, some of the science materials provided ideas for situations on which questions were based (e. g. Hedges, 1960; Klopfer, 1964).

Summary of Procedures for Test-item Construction: The Nay-Crocker inventory of affective attributes of scientists (Appendix A) was used as a framework for general affective objectives for science education, and the present study examines eight of these attributes which are listed in the inventory under the heading of attitudes. A list of behaviors stated in the context of the science classroom was compiled to define each of the attributes. These lists were reduced on the basis of the responses of a panel of judges who indicated whether or not they felt thet each behavior was relevant to the attribute under which it was listed.

Multiple-choice questions (TOSA) to reflect the defining behaviors were then written. The stem of each question describes a situation and the four alternate responses describe courses of action which could be taken in relation the situation. Each keyed response was designed to be consistent with one of the behaviors defining the attributes. Initially, attempts were made to write questions in which all four alternatives were related to the same behavior, but for most questions this was not possible. For some questions, the keyed response is related to one of the defining behaviors, while some of the other alternatives are related to different behaviors.

The discussion in this chapter up to this point has dealt with the rationale and procedures for the construction of the test items and a description of the test items which were constructed. The remainder of this chapter deals with the procedures for data collection and analysis.

II. Data Collection

Population: The population from which the sample was drawn consists of Chemistry 20 and Physics 20 classes in the Edmonton Public School System.

Sample: Selected Chemistry 20 and Physics 20 classes from two schools in the Edmonton Public School System participated in this study. Students from a third school participated in the pilot testing of the first draft of 28 test items. The number of students involved at each testing is given in the discussion of procedures.

Measuring Instruments: The Test On Scientific Attitude (TOSA), the Test Of Likert Items (TOLI), and the Watson-Glaser Critical Thinking Appraisal (Form Ym) were administered to samples of students as a part of this study. Student scores on the Cooperative School and College Ability Test (SCAT, form 3A or 3B) and the the Cooperative Sequential Test of Educational Progress in reading (STEP reading, form 3A or 3B) were obtained from the grade 9 records at the division of testing and research of the Department of Education in Alberta.

1. TOSA: This is a forty-item, multiple-choice test which was constructed as a part of this study (see Appendix B). The test has two subtests each of which is 20 items long. The test content has already been discussed in this chapter under the description of the test format. Item 20 was not included in the data analysis because of a typing error which ocurred in the test which was administered to the student sample. The test items are scored 1-0 with one keyed response for each question. A panel of judges was used to confirm the selected keyed responses.

2. TOLI: This test consists of twenty-five opinion statements relevant to the scientific attributes which are being examined in the present study. The students are asked

to respond to each statement on a Likert scale consisting of the following four response categories: strongly agree, partly agree, partly disagree and strongly disagree. Some of the items were selected from the science-attitude scales discussed in Chapter II (see Appendix C). The remaining statements were written for the purpose of this study. The recommendations made by Likert (1932) and Oppenheim (1966) for the writing of opinion statements were followed in the selection of statements for this test.

Traditional scoring of the Likert scale used in this test would assign a value of 4 for strong agreement with a positive statement and strong disagreement with a negative statement. The remaining three responses would be assigned values of 3, 2, and 1. However, it was felt that, for some of the statements, partly agree or partly disagree were more consistent with the attitudes being measured. Statement 14, "When something is explained well, there is no reason to look for another explanation", is an example of such a statement.

The response which was assigned a value of 4 for each of the statements is underlined in Appendix C. If PA is assigned a value of 4, then SA, PD, and SD are assigned values of 3, 2, and 1 respectively. This system was applied to all the statements in this test. The response which was assigned the value of 4 was selected on the basis of the responses of a panel of judges.

3. Watson-Glaser Critical Thinking Appraisal: This test is designed to measure five aspects of critical thinking ability: "ability to discriminate among degrees of truth or falsity of inferences drawn from given data", "ability to recognize unstated assumptions", "ability to reason deductively", "ability to weigh evidence", and "ability to distinguish between arguments which are strong and relevant and those which are weak or irrelevant" (Watson and Glaser, 1964, p. 2).

The odd-even, split-half reliability coefficient corrected by the Spearman-brown formula is 0.86 for a sample of 2406 grade 11 students. The corresponding coefficient for various other groups ranges from 0.85 to 0.87. The test has an average correlation coefficient of 0.73 with the Otis Mental Ability Tests: Gamma for a sample of 20,312 grade 9 to grade 12 students and a correlation coefficient of 0.66 with the STEP reading test for a sample of 318 grade nine students. The authors present convincing arguments for content and construct validity. The Seventh Mental Measurements Yearbook (Buros, p. 783) cites 109 studies in which this test has been used.

4. SCAT: This test consists of a fifty-item verbal subtest and a sixty-item quantitative subtest. The verbal subtest measures the ability to comprehend the sense of a sentence and the ability to attach meanings to isolated words. The quantitative subtest measures the ability to

manipulate numbers, the ability to apply number concepts in computational situations and the ability to solve quantitative problems (Cooperative Test Division of the Educational Testing Service, 1958). Forms 3A and 3B have a difficulty level appropriate for grade 9 students. These two forms have been shown to be equivalent in that raw scores from these two forms give equivalent scores when converted to the same standardized scale.

The KR-20 reliability coefficients based on a sample of 2880 grade 9 students is 0.93 for the verbal subtest, 0.89 for the quantitative subtest, and 0.95 for the total test. The total score, rather than the score on either subtest, was found to be the most reasonable predictor for science achievement. The predictive validity for science achievement over a two year period (from grade 9 to 11) is 0.43. The average correlation between SCAT scores and science achievement scores is 0.63.

5. STEP Reading: This test measures the ability to understand direct statements, the ability to interpret and summarize a passage and the ability to see the motives of the author (Cooperative Test Division of the Educational Testing Service, 1957). Forms 3A and 3B are appropriate for use at the grade 9 level. These two forms have been shown to be equivalent in that raw scores from these two forms give equivalent scores when converted to the same standardized scale and the two forms have similar

distributions of item difficulties.

The KR-20 reliability coefficient based on a sample of 408 grade 8 students is 0.90. The average correlation between total SCAT scores and STEP reading scores is 0.81 for samples of grade 9 students ranging in size from 200 to 225.

Procedures: The abbreviations for the test names indicated under the discussion of testing instruments above, are used throughout the present section.

The procedures followed to obtain behavioral definitions of the affective attributes and in the construction of test items for TOSA have been discussed in the present chapter under the rationale.

The first draft of 28 test items was administered to a sample of 76 Chemistry 20 and Physics 20 students from one school in the Edmonton Public School System. Students from this school did not write the final draft of the test. The results of item analysis, and student comments on the reading level of the test items, possible ambiguity of statements, and reasons for selecting various responses were used to revise some of these items. The information obtained from this pilot run was also used to guide the construction of another twelve items to made up the 40 items in Appendix B.

During the spring semester, TOSA and TOLI were administered to a sample of three classes of Chemistry 20 students and three classes of Physics 20 students from two schools in the Edmonton Public School System. 156 students were tested. 132 of these students also wrote the Watson-Glaser Critical Thinking Appraisal. The teachers of these classes provided a rating for each student to indicate the extent to which the student exhibited the behaviors defining the affective attributes. This rating is on a four-point scale for which a rating of 4 indicates frequent demonstration of these behaviors. Only one rating on the overall set of behaviors for the six attributes was obtained for each student. The instructions to the teachers are given in Appendix E.

SCAT (form 3B) and STEP reading (form 3B) scores for 118 students of this sample were obtained from the 1970 grade 9 records at the division of testing and research of the Department of Education. The reading score is expressed as a percentile based on the total population of grade 9 students who wrote this test in 1970. For SCAT, raw scores on the verbal subtest (out of 60) and on the quantitative subtest (out of 50) were obtained. These were added together to provide a total score.

During the following fall semester, TOSA was administered to a sample of 4 classes of Chemistry 20 students and 3 classes of Physics 20 students from the same two schools (130 students). Three weeks following the first administration, this same sample of classes wrote TOSA a second time (126 students). The purpose of the second testing was to provide data for the calculation of a test-retest correlation coefficients. This second sample of students was necessitated because time did not permit a retest during the spring semester. 151 different students wrote the test during these two administrations and 105 students wrote the test twice. SCAT (form 3A) and STEP reading (form 3A) scores for 134 students of this sample were obtained from the 1971 grade 9 records at the Department of Education. Teacher ratings for these students were not obtained.

III. Data Analysis

The procedures followed in the analysis of the data and the reasons for the inclusion of each step of the analysis are discussed in this section. The statistical tests and other analytic techniques which are used are also described.

External (concurrent) Validity: The high student group and the low student group (see definition of terms in Chapter I) are used to examine the external validity of TOSA, the two subtests of TOSA, and TOLI. Since these groups were established on the basis of the teacher ratings, only students from the sample of 156 students who wrote during the spring semester are included in this analysis. Claims for concurrent validity will be made if it can be shown that the high student group scores significantly higher than the low student group on the criterion measures (TOSA, its two subtests, and TOLI), that is, if the null hypothesis (Hypothesis I in Chapter I) can be rejected for the various criterion measures. Hypothesis I was tested by a one-way analysis of covariance with two factor levels (high student group and low student group). A separate analysis was performed for each of the four criterion measures. General scholastic ability as measured by total SCAT scores is the covariate in these analyses. Correlations of test scores with teacher ratings will also be reported.

Analysis of covariance, rather than analysis of variance was used because the students were not randomly assigned to the two groups. In this situation, analysis of covariance can be employed to remove potential biases in assigning students to groups. (Winer, 1962, p. 578). In the case of the present study, the teacher's rating may have been influenced by the student's scholastic ability. In analysis of covariance, student scores are adjusted to account for any difference in scholastic ability which may exist between the two groups.

The use of analysis of covariance in the present study as described above may be somewhat questionable. Even if students could be ideally divided into a high and low group based solely on the characteristics which TOSA is designed to measure, one might expect the high student group to have a higher general ability than the low student group. However, if it can be demonstrated that the scores on TOSA are significantly different after the effect of general ability is removed, the claim for the concurrent validity of the test will be that much stronger.

There are a number of assumptions underlying the application of analysis of covariance. The assumptions underlying analysis of variance also apply to the analysis of covariance, that is, the normal distribution of scores and the homogeneity of variance among groups. The additional assumptions of linear regression (homogeneity of residual variance) and homogeneity of regression among groups apply to analysis of covariance. The use of the F-test in the analysis of covariance is robust with respect to violation of the normality assumption and the assumption of homogeneity of residual variance (Winer, 1962, p. 586). In the present study a test is made to demonstrate the homogeneity of within-group regression. The computational procedures involved in the analysis of covariance are given in Winer (1962, pp. 581-594). The ANCV10 computer program provided by the Division of Educational Research Services at the University of Alberta was used to do the calculations.

Differences Between Samples: The statistical technique of t-tests between independent samples was used to test for differences in scholastic ability, reading ability, and TOSA scores between the students who wrote during the spring and the students who wrote during the fall (test of Hypothesis III in Chapter I). The results of this analysis were used to decide whether or not the two samples should be combined for the item analysis and the factor analysis described below. A larger sample will provide more stable correlations for the factor analysis and the use of a large sample will decrease the sampling error associated with the biserial correlations calculated in the item analysis.

The computational procedures involved in this analysis are given in Winer (1962, pp. 14-36). The test for homogeneity of variance was also made. The ANOV10 computer program provided by the Division of Educational Research Services was used to do the calculations.

Item Analysis: The TEST04 program provided by the Division of Educational Research Services was used to obtain the following information for the test-items in TOSA and its subtests: percentage of students selecting each alternative, biserial correlations, KR-20 coefficients, and total-score distributions.

McNemar (1949, pp. 215-221) discusses the use of biserial correlation coefficients to describe relationships between dichotomous and continuous variables. The biserial coefficient rather than the point-biserial coefficient is used when it can be assumed that there is a normally-distributed continuous variable underlying the dichotomy. The assumption of linear regression is also made. McNemar indicates that the main issue to be considered is the question of continuity. This can be argued on the basis of the nature of the characteristic being measured. In the present study, it is not likely that all those students selecting the keyed response are at the same level with respect to the attribute to which the question is related.

Although the biserial correlation coefficient is theoretically free from bias toward extremely easy or difficult items (Gulliksen, 1950, p. 393), the sampling error associated with this coefficient is quite large if the dichotomies are extreme (McNemar, 1949, p. 217). This sampling error can be reduced somewhat by the use of large sample sizes.

The alpha reliability coefficient and item-to-total correlation coefficients were obtained for TOLI. The DEST01 and DEST02 computer programs provided by the Division of Educational Research Services were used for the calculations.

Factor Analysis: Factor analysis has been designed to identify a set of underlying or latent variables (smaller in number than the original set of observed variables) which can maximally reproduce the correlations between the observed variables (Harman, 1960, p. 15). A factor loading matrix, in which each variable has a loading on each factor, is obtained. These variable loadings are regression coefficients for predicting variable scores from factor scores. If r factors are retained, these r factors represent the orthogonal axes in an r-dimensional space and the loadings represent the projections of the variables on these axes. These axes are rotated by a transformation on the factor loading matrix to give a factor pattern matrix which ideally has a few large loadings and a large number of near-zero loadings. The group of variables which have high loadings on the same factor will be that set of variables which are positioned close together in the r-dimensional space. That is, those variables which are most highly correlated with each other.

The common factor model was used in the present study. In this model each variable is considered to be composed of a common and a unique part. The common part of each variable is that portion of its variance that it has in common with the other variables in the domain of interest. The communality of the variables is the squared multiple correlation coefficient of that variable with all of the other variables in the domain of interest. Since data is not available for the complete set of variables in the domain, an estimate, rather than the exact value, of the communality must be obtained. This estimate can be obtained by selecting an initial estimate (e. g. the squared multiple correlation of the variable with the other n-1 variables in the study) and then revising this estimate through iterative procedures (Harman, 1960, pp. 68-92).

In the present study, an unweighted least squares method of factoring was used to obtain the factor loading matrix. This procedure utilizes a roots and vectors decomposition of the $R-U^2$ matrix and the final solution is determined by minimizing one-half the trace of $(R-R^*)^2$ where R is the observed correlation martix and R* is the reproduced correlation matrix (Hakstian and Bay, 1972, p. 21). The $R-U^2$ matrix is the covariance matrix of the common parts of the variables. The off-diagonal elements are the same as the off-diagonal elements of R and the diagonal elements are the communalities of the variables. Since the test-item scores are dichotomous scores for which an underlying continuum can be assumed, the tetrachoric correlation matrix was used. These correlations were calculated by the cosine-pi formula, and are therefore biased in the case of test items with extreme difficulty levels (Ferguson, 1959, p. 244).

Since the dimensions of the Test On Scientific Attitude are not expected to be uncorrelated, the factor loading matrix was rotated to an oblique factor pattern matrix. The rotational procedure outlined by Harris and Kaiser (1969) was used. The decision to apply this method was made on the basis of research which indicates that the Harris-Kaiser rotational procedure, when compared with other oblique rotational procedures, consistently gives solutions which more closely approximate simple structures (Hakstian, 1971).

The Alberta General Factor Analysis Program was used to do the computation (Hakstian and Bay, 1972).

Other Test Statistics: The results for the sample of 105 students who were present for both administrations of the test during the fall semester were used to provide test-retest correlations for TOSA and its two subtests.

Test means, standard deviations, and distributions were obtained. Correlations between scores on TOSA, its two subtests, and TOLI and scores on SCAT, STEP reading and the Watson Glaser Critical Thinking Appraisal were obtained. The DEST05 computer program provided by the Division of Educational Research Services was used in the calculation of these correlations.

Hypothesis II stated in Chapter I (differences between sexes) was tested by a one-way analysis of covariance in which the two groups are male and female students and the covariates are general scholastic ability as measured by SCAT, and reading ability as measured by STEP.

Chapter IV

Results And Discussion

The results of the present study are presented and discussed in this chapter in five sections. Arguments relative to content validity are presented in the first section. The second section includes the statistical tests for Hypothesis II and Hypothesis III stated in Chapter I, and a number of general statistics such as means, standard deviations, and correlations. The results relevant to structural validity are given in the third section. These are the results of the item analysis and the factor analysis. Test stability is reported in the fourth section. The tests of Hypothesis I are presented in the fifth section under the heading of external validity. The correlations of the test scores with teacher ratings are also reported in this section.

The following abbreviations are used throughout this chapter to refer to the tests which are described in Chapter III:

TOSA - Test On Scientific Attitude
CCS - Cognitive Component Subtest of TOSA
ICS - Intent Component Subtest of TOSA
TOLI - Test Of Likert Items
SCAT - Cooperative School And College Ability Test
STEP - Cooperative Sequential Test of Educational Progress in reading
WCTA - Watson-Glaser Critical Thinking Appraisal.

TOSA, CCS, ICS, TOLI, and WCTA scores are expressed in percent. STEP scores are percentiles and the verbal, quantitative, and total SCAT scores are raw scores out of 60, 50, and 110 respectively.

I. Content Validity

The content validity of the test items constructed for this study (Appendix B) can be argued on the basis of the procedures that were followed in the selection of the attitudes to be measured, in defining these attitdes, and in the construction of the items.

Defining the Dimensions of the Test: The six attitudes which the test items in TOSA were designed to measure were selected from a list of affective attributes which Nay and Crocker (1970) feel should be demonstrated by scientists and science students. This list of attributes was compiled on the basis of interviews with scientists and a survey of the literature related to the nature and philosophy of science.

The results of the panel ratings which were used to select behavioral definitions for these attributes are presented and discussed in Chapter III. Test items were constructed to reflect only those behaviors which the panel indicated to be important defining characteristics of the attributes.

Item Content And Scoring Key: The test items describe science-related situations in which the defining behaviors could be exhibited. A wide variety of science reading materials were surveyed in search of ideas for behavioral specification of the attributes and science related situations on which to base the test items (e. g. Diederich, 1967; Eiss and Harbeck, 1969; Hedges, 1966; Klopfer, 1964).

Three experienced science teachers, two of whom were working towards a Ph.D. in secondary science education, examined the final draft of the test questions to provide a scoring key. Question 30 was the only question on which more than one member of this panel disagreed with the keyed response proposed by the two people responsible for constructing the test items. For questions 4, 19, 22, 27, 35 and 36, one disagreement with the proposed key was recorded. All but three of these disagreements were resolved through a discussion of the intentions of these test items. The disagreements for the following questions were not resolved:

22. Imagine that you have just finished a laboratory investigation. Your measurements all agree except two. Which of the following would you do?

    A. Include the two odd measurements in your report but omit them from calculations.
    B. Adjust the two odd measurements to make them agree better with the others.
    C. Take more measurements.
    D. Use all the measurements as they are when making calculations.

35. In an experiment, students blew through limewater and noted that it turned milky. From this result, most of them concluded that their bodies give off carbon dioxide. However, one girl wrote in her notebook that since there is carbon dioxide in the air we breathe, the experiment proved nothing. Which one of the following best describes your evaluation of this statement?

    A. The students were justified in making their

conclusion.
B. The girl was justified in doubting the proof.
C. Neither side had sufficient grounds for their statements.
D. Both sides were partly justified in their statements.

36. "People born when certain stars are becoming more prominent show the influence of these stars in their personalities." People who believe this statement

A. probably have a special ability to understand such influences.
B. are not critical enough.
C. are more openminded than most people.
D. have a disregard for scientific evidence.

The accepted keyed response to the above questions are underlined. One panel member felt that the keyed response to question 22 should be B. The item difficulty for this question is 0.11. The average score on TOSA for those students who selected A is 20.7 out of 39 as compared with average scores of 20.6, 20.1 and 16.6 for those students who selected alternatives D, C, and B respectively. The biserial correlation for this question with the total test (39 questions) is 0.007 indicating that this question is not closely related to the other questions in the test. This question was designed to measure honesty in reporting results. However, responses C and D may be acceptable answers depending on the number of observations which were taken and the extent of the disagreement of the observations. The item analysis indicates that this question should be revised. A possible revision would be to provide more information and to make all the alternatives more closely related to reporting of results.

One panel member felt quite strongly that the keyed response to question 35 should be B. He objected to students drawing their conclusion solely on the basis of this experiment. However, the author of this research study feels that the presence of the word "partly" in alternative D makes this an appropriate keyed response. This question was designed to measure the suspended judgement in interpreting experimental results (generalizes only to the degree justified by available evidence). The difficulty for this question is 0.50. The average score on TOSA for those students who selected D is 21.4 out of 39 as compared with an average score of 19.6, 19.0, and 18.0 for those students who selected alternatives B, C and A respectively. The biserial correlation for this question is 0.35. These item statistics indicate that this is an acceptable test item.

One panel member felt that alternatives B and D were equally acceptable responses for question 36. The author of the present study feels that B is a more acceptable response because there is not a great deal of empirical evidence to contradict this belief. The difficulty for this question is 0.18. The average score on TOSA for those students who selected alternative B is 21.8 out of 39 as compared with average scores of 20.2, 19.8 and 18.9 for those students who selected alternatives D, C, and A respectively. The biserial correlation for this question is 0.25. Although the difficulty is quite low, the other statistics for this item are acceptable. This item was designed to measure critical mindedness (challenges the validity of unsupported statements).

The panel described above also responded to the statements in TOLI. For statements 9, 10, and 11, all three panel members disagreed with the keyed response proposed by the two people responsible for writing and compiling the statements in this test. The panel members responded partly agree to these three statements while the proposed keyed response was strongly agree.

One disagreement was recorded for all of the statements except 2, 6, 7, 13, 17, and 19. Most of these disagreements were recorded by one panel member who showed a strong tendency to respond PA or PD. He gave one of these responses to eighteen of the twenty-five statements. The maximum number of PA or PD responses for any one of the other four people involved was eight. It should be noted that this same panel member agreed with all but two of the proposed keyed responses for the forty questions in TOSA.

Summary: The content validity of the test items has been argued on the basis that the attitudes which the test is designed to measure were selected from a list of affective attributes of scientists, the behavioral specification of these attributes were selected on the basis of the responses of a panel of judges, the content of the items describe science-related situations, and the content of the items is comparable with the ideas expressed in a wide variety of science reading materials. The validity of the keyed responses has also been demonstrated by a panel of judges.

II. Descriptive Statistics

The tests for Hypothesis II and Hypothesis III stated in Chapter I and a number of means, standard deviations, and correlation coefficients are presented in this section.

Tests for Differences Between Samples: Two samples participated in the present study. Those 156 students who were tested during the spring semester will be identified as GROUP 1. Those 151 students who wrote during the fall semester will be identified as GROUP 2. The information obtained for these samples is described in Chapter III in the section on procedures. Independent-sample t tests were used to test the following null hypotheses:

Hypothesis IIIa: There is no significant difference between the mean score on SCAT for GROUP 1 and the mean score on SCAT for GROUP 2.

Hypothesis IIIb: The above hypothesis stated for STEP scores.

Hypothesis IIIc: The above hypothesis stated for TOSA scores.

If the above hypotheses are not rejected, the test results for the two samples will be combined for data analysis dealing with test results which are available for both groups. Table I gives the results of the F tests for differences between variances and Table II gives the results of the t tests for differences between means.

Table I.

F TESTS FOR VARIANCE DIFFERENCES BETWEEN GROUP1 AND GROUP 2

TEST	VARIANCE		DF		F	P
	GRP1	GRP2	GRP1	GRP2		
SCAT	182.7	209.0	117	129	1.14	0.46
STEP	201.9	196.1	117	129	1.03	0.87
TOSA	106.5	107.4	155	150	1.01	0.96

Table II.

T TESTS FOR MEAN DIFFERENCES BETWEEN GROUP1 AND GROUP 2

TEST	MEAN		DF	T	P
	GRP1	GRP2			
SCAT	77.2	78.7	246	0.82	0.41
STEP	65.2	64.2	246	0.56	0.57
TOSA	53.2	51.6	305	1.35	0.18

Conclusions: Hypotheses IIIa, IIIb and IIIc are not rejected. The mean scores on SCAT, STEP, and TOSA for GROUP1 are not significantly different from the mean scores for GROUP 2. The variances of the GROUP 1 scores on SCAT, STEP, and TOSA are not significantly different from the variances of the GROUP 2 scores. These two groups will be treated as one sample of 307 students for the calculations of some of the statistics reported below, and for the item and factor analysis of the test items in TOSA.

The use of the data for this larger sample for the item analysis will reduce the sampling error associated with the biserial correlation coefficients for items which have extreme difficulty levels (McNemar, 1949, p. 217), and will give more stable correlation coefficients for the factor analysis and for comparing the scores on TOSA with other test scores. Ebel (1965, p. 273) reports evidence which indicates that correlation coefficients for sample sizes of 300 are considerably more stable than those for sample sizes of 100. This is particularly true for coefficients which are lower than 0.6. The correlation coefficients reported in Table V are based on the combined groups where data are available for both GROUP1 and GROUP 2.

The SCAT and STEP scores obtained for the students in GROUP 1 are scores on form 3B tests while form 3A test scores were obtained for the students in GROUP 2. However, these two forms of the SCAT and STEP tests have been shown to be equivalent (see the discription of these tests under measuring instruments in Chapter III). The correlations between total SCAT and STEP scores are 0.64 for GROUP 1 (118 students), 0.60 for GROUP 2 (130 students) and 0.62 for the combined group. These correlation coefficients indicate that the scores obtained for these two groups are comparable.

Means, Standard Deviations, Ranges, and Distributions: A summary of the test means, standard deviations, and ranges of total scores is given in Table III. All test scores are reported in percent except STEP scores which are percentiles, and SCAT scores which are raw scores. The totals possible for SCAT are 60 for the verbal subtest (SCAT-V), 50 for the quantitative subtest (SCAT-Q) and 110 for the total test (SCAT-T). Scores for the Watson-Glaser Critical Thinking Appraisal (WCTA) and the Test Of Likert Items (TOLI) were obtained only for those students who wrote during the spring semester. The statistics for all the other tests are reported for the combined sample.

Table III.

TEST MEANS, STANDARD DEVIATIONS, AND RANGES

TEST	N	MEAN	STD. DEV.	MIN.	MAX.
TOSA	307	52.4	10.3	21	79
CCS	307	52.8	13.9	16	84
ICS	307	52.2	12.5	15	90
TOLI	156	83.2	6.0	65	96
SCAT-V	248	43.2	8.5	18	59
SCAT-Q	248	34.8	7.5	15	49
SCAT-T	248	78.0	14.0	30	115
STEP	248	64.7	14.1	30	96
WCTA	132	68.4	8.8	49	86

The distributions of scores on TOSA, CCS, ICS, and TOLI are given in Table IV. The distribution for ICS appears to be the closest approximation to a normal distribution. The manner in which the Likert items are scored contributes to the tendency for the scores on this test to cluster about the 80 percent level. All responses that a student makes contribute to his total score on TOLI.

Table IV.

FREQUENCY DISTRIBUTIONS OF TEST SCORES

	FREQUENCIES			
INTERVAL	TOSA	CCS	ICS	TOLI
10-19		5	2	
20-29	6	9	5	
30-39	28	38	30	
40-49	81	72	69	
50-59	127	97	96	
60-69	52	53	75	3
70-79	13	30	27	34
80-89		3	2	98
90-99				21
TOTAL	307	307	307	156

Correlations: The correlation coefficients for the combined sample of 307 students are given in Table V. This table also includes the probabilities that r=0. Scores for the WCTA and TOLI were obtained only for those students who wrote during the spring semester. The correlations between these two tests and the other tests are given in Table VI.

Table V.

CORRELATIONS FOR THE SAMPLE OF 307 STUDENTS
THE UPPER TRIANGLE CONTAINS THE PROBABILITIES THAT r=0

TEST	TOSA	CCS	ICS	STEP	SCAT-V	SCAT-Q	SCAT-T
TOSA	1.00	0.00	0.00	0.00	0.00	0.11	0.00
CCS	0.80	1.00	0.00	0.00	0.00	0.00	0.00
ICS	0.77	0.23	1.00	0.05	0.02	0.61	0.26
STEP	0.35	0.41	0.13	1.00	0.00	0.00	0.00
SCAT-V	0.38	0.46	0.15	0.64	1.00	0.00	0.00
SCAT-Q	0.18	0.31	-0.03	0.42	0.52	1.00	0.00
SCAT-T	0.33	0.44	0.07	0.62	0.89	0.85	1.00

N=307 for the correlations in the first three rows.
N=248 for the correlations in the last four rows.

Table VI.

CORRELATIONS FOR GROUP 1
THE SECOND ROW CONTAINS THE PROBABILITY THAT r=0
THE THIRD ROW CONTAINS THE SAMPLE SIZE

TEST		TOSA	CCS	ICS	STEP	SCAT-V	SCAT-Q	SCAT-T	WCTA	TOLI
TOLI	r	0.37	0.37	0.24	0.27	0.35	0.16	0.30	0.36	1.00
	p	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00
	N	156	156	156	118	118	118	118	132	156
WCTA	r	0.41	0.45	0.24	0.57	0.59	0.49	0.62	1.00	0.36
	p	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	N	132	132	132	107	107	107	107	132	132

The correlation of 0.23 between the two subtests of TOSA (CCS and ICS) tends to indicate that these two subtests are not measuring the same characteristics. This correlation is considerably lower than the odd-even, split-half correlation for TOSA (0.40).

The tests constructed for the present study (TOSA, CCS, ICS, and TOLI) have fairly low correlations with reading ability and general scholastic ability. In fact, ICS scores have a zero correlation with both the quantitative and total SCAT scores. The correlations of these tests with the Watson-Glaser Critical Thinking Appraisal are also quite low. Again, ICS has the lowest correlation coefficient. High correlations were not expected because the Critical Thinking Appraisal was designed to measure abilities (see the description of this test in Chapter III) while the tests constructed for this study were designed to measure attitudes.

The Test Of Likert Items was included in the present study to provide a comparison between test items of the Likert format and the multiple-choice items in TOSA. The correlation between TOLI and TOSA (0.37) is surprisingly low since statements were included in TOLI only if their content was similar to the content of the questions in TOSA. A possible reason why the two tests are not more highly correlated may be that each statement in TOLI is considered separately from the others while each alternative in TOSA is considered in relation to the situation described in the stem and in relation to the other distractors. It is also possible that the response biases associated with Likert items (see Chapter II) may have had some influence on the student scores on TOLI.

Tests for Sex Differences: A one-way analysis of covariance in which the factor levels are males and females and the covariates are scholastic ability as measured by SCAT and reading ability as measured by STEP was used to test Hypotheses IIa, IIb, and IIc. Both reading ability and scholastic ability are used as covariates because the males in this sample have a higher mean SCAT score but a lower mean STEP score. The null hypotheses are stated below:

Hypothesis IIa: There is no significant difference between the mean score on TOSA for male students and the mean score on TOSA for female students when scholastic ability and reading ability are the covariates.

Hypothesis IIb: The above hypothesis stated for CCS scores.

Hypothesis IIc: The above hypothesis stated for ICS scores.

The results of these analyses are given in Tables VII, VIII and IX.

Table VII.

ANALYSIS OF COVARIANCE ON TOSA SCORES MALES VS FEMALES

SOURCE	DF	MEAN SQUARE	F	P
Between Groups	1	4.25	0.04	0.83
Error	244	95.9		

	SAMPLE SIZE	GROUP MEANS			
		UNADJ	ADJ	SCAT	STEP
MALES	157	51.9	52.0	78.7	62.7
FEMALES	91	52.6	52.3	76.6	68.1

Table VIII.

ANALYSIS OF COVARIANCE ON CCS SCORES MALES VS FEMALES

SOURCE	DF	MEAN SQUARE	F	P
Between Groups	1	278.8	1.82	0.18
Error	244	153.0		

	SAMPLE SIZE	GROUP MEANS			
		UNADJ	ADJ	SCAT	STEP
MALES	157	51.4	51.5	78.7	62.7
FEMALES	91	53.9	53.8	76.6	68.1

Table IX.

ANALYSIS OF COVARIANCE ON ICS SCORES
MALES VS FEMALES

SOURCE	DF	MEAN SQUARE	F	P
Between Groups	1	131.8	0.83	0.36
Error	244	159.5		

		GROUP MEANS			
	SAMPLE SIZE	UNADJ	ADJ	SCAT	STEP
MALES	157	52.3	52.6	78.7	62.7
FEMALES	91	51.4	51.0	76.6	68.1

In the above analyses, the probabilities that the assumption of homogeneity of regression is satisfied are 0.54, 0.61, and 0.69 for TOSA, CCS, and ICS respectively.

Conclusions: Hypotheses IIa, IIb, and IIc are not rejected. The mean scores for female students on TOSA, CCS, and ICS are not significantly different from the mean scores for male students. The adjusted means and the unadjusted means do not differ very much because the effect of the two covariates are in opposite directions.

III. Structural Validity

Two aspects of the structural validity of the test items constructed for this study are examined in this section. The properties of the individual items and the test-homogeneity are discussed in the light of the results of item analysis. Factor analysis is used to examine the underlying structure of the test items in TOSA.

Item Analysis: Item analysis was performed on TOSA, its two subtests, and TOLI.

1. TOSA: The percentage of students out of 307 who chose each response is given in Appendix B. Students were allowed sufficient time to complete the test and all of the students responded to all of the questions. Most students required between twenty-five and thirty minutes to write TOSA. Most of the alternatives proved to be acceptable distractors. Out of the 156 alternatives analyzed for the thirty-nine questions, seven alternatives received less than three percent of the responses and none of the distractors were completely ignored. Most of the distractors that received one or two percent of the responses are in questions with difficulty levels of 0.80 or higher. It is possible that the nature of the sample which was tested may have contributed to raising the difficulty level of some of the easier questions. Chemistry 20 and Physics 20 are not compulsory courses. Therefore, most of the students registered in these courses will be taking the course because of their interest in the subject. Also, a certain amount of screening is done before students are allowed to take these courses.

The item difficulties and biserial correlation coefficients are summarized in Table X. Question 20 was omitted from the analysis because of a typing error which occurred in the test copy administered to the students.

Biserial correlations for each test item with the total test score (TOSA) and with the two subtest scores (CCS and ICS) are given in Table X.

The item difficulties range from 0.11 for question 22 to 0.87 for question 26. Twenty-seven items have an item difficulty level between 0.25 and 0.75. Seven have item difficulty levels higher then 0.75 and five have item difficulty levels lower than 0.25.

Questions 19 and 22 have a zero biserial correlation coefficient with TOSA and questions 1, 8, 11, 38, 39, and 40 all have coefficients less than 0.25. The biserial correlation for item 22 with ICS is 0.24. This is a considerable increase over 0.07, but is still quite low since there are only 19 items in this subtest. The biserial correlation for item 19 does not show any significant increase when it is calculated for CCS. The biserial correlations for questions 8, 14, 25, 38, and 40 increase considerably when calculated for the subtest to which they belong. The biserial correlations for questions 1, 11, and 39 increase only slightly when calculated for the subtests. The biserial correlations of the items with the subtest scores will be spuriously high because of the small numbers of items in these subtests. However, most of them are above 0.30 and should be satisfactory.

Table X.

DIFFICULTY LEVELS AND BISERIAL CORRELATIONS FOR TEST ITEMS IN TOSA

ITEM	DIFFICULTY	BISERIAL CORRELATIONS		
		TOSA	CCS	ICS
1	0.85	0.22	0.29	
2	0.61	0.37	0.36	
3	0.59	0.34	0.49	
4	0.52	0.27	0.27	
5	0.66	0.40	0.46	
6	0.55	0.44	0.46	
7	0.52	0.47	0.54	
8	0.32	0.23	0.37	
9	0.82	0.41	0.43	
10	0.80	0.60	0.60	
11	0.54	0.22	0.25	
12	0.55	0.33	0.42	
13	0.22	0.29	0.41	
14	0.40	0.21	0.35	
15	0.26	0.25	0.38	
16	0.50	0.30	0.37	
17	0.58	0.42	0.57	
18	0.42	0.41	0.52	
19	0.18	0.01	0.08	
21	0.20	0.24		0.32
22	0.11	0.07		0.24
23	0.37	0.27		0.33
24	0.64	0.27		0.35
25	0.35	0.21		0.30
26	0.87	0.65		0.63
27	0.58	0.32		0.38
28	0.68	0.36		0.47
29	0.60	0.38		0.46
30	0.49	0.35		0.38
31	0.81	0.34		0.51
32	0.86	0.48		0.62
33	0.85	0.51		0.44
34	0.39	0.37		0.41
35	0.50	0.34		0.38
36	0.18	0.25		0.29
37	0.39	0.27		0.41
38	0.56	0.19		0.33
39	0.56	0.20		0.27
40	0.34	0.15		0.34

Examination of question 19 which is given below indicates that it may be a trick question, and the reasons for choosing the keyed response may have little in common with the attitudes being measured.

> Which one of the following is NOT an important reason why scientists often repeat experiments reported by other scientists?
>
> A. A scientist could be so intent on finding a specific answer that he might subconsciously observe only what he wants to see in his experiments.
> B. This helps to keep scientists careful and honest when making observations and reporting results.
> C. Other scientists might give a different interpretation to the same observations.
> D. The first scientist might overlook a significant variable in his experiment.

C is the keyed response because giving different interpretations to the same observations, is not a reason for repeating the experiment. The second scientist could examine the results reported by the first scientist.

This question could possibly be changed to be more consistent with the other questions by replacing alternative C with a different distractor and using alternative B as the keyed response. The question then would be consistent with the behaviors which define honesty. That is, scientists should not require this type of checking to demonstrate honesty in reporting results.

Question 22 which is given below has two apparent weaknesses.

> Imagine that you have just finished a laboratory investigation. Your measurements all agree except

two. Which one of the following would you do?

A. Include the two odd measurements in your report but omit them from calculations.

B. Adjust the two odd measurements to make them agree better with the others.

C. Take more measurements.

D. Use all the measurements as they are when doing calculations.

The most appropriate alternative would depend on the number of original observations taken and on the degree of the discrepancy. This item appears to be somewhat process oriented. The situation described in this question is related to honesty in reporting data, but the wording of the stem and the distractors to be used must be changed to be more consistent with the intent of the item.

2. TOLI: This test was administered only to the 156 students who wrote during the spring semester. The percentage of students selecting each alternative is given in Appendix C. Out of the twenty-five questions in this test, two alternatives received no responses, six alternatives received one percent of the responses, and three alternatives received two percent of the responses. The percentage of students selecting those responses that were assigned a scale value of 4 range from 31 to 93. On the Test Of Likert Items, a greater proportion of the alternatives received few responses than on the multiple-choice test.

Table XI gives a list of the product moment correlations between the test items and the total test score.

Table XI.

ITEM TO TOTAL CORRELATIONS FOR TOLI

ITEM	CORRELATION	ITEM	CORRELATION	ITEM	CORRELATION
1	0.20	10	0.16	18	0.17
2	0.31	11	0.21	19	0.38
3	0.38	12	0.30	20	0.38
4	0.34	13	0.37	21	0.19
5	0.36	14	0.27	22	0.48
6	0.27	15	0.38	23	0.14
7	0.31	16	0.42	24	0.26
8	0.23	17	0.44	25	0.30
9	0.37				

The probability that the correlation coefficient for item 23 is zero is 0.09. The other correlations are all significantly different from zero. Following is a list of those statemensts with low correlations with the total test score:

1. When a scientist is shown enough evidence that one of his ideas is a poor one, he should change it.
10. Many ideas which scientists find to be useful may not be entirely correct.
11. It is necessary to question periodically the basic truths of science.
18. A person should not make up his mind until he has collected as many facts as possible.
21. Once a person makes up his mind he should be reluctant to change it.
23. When making decisions about drinking alcohol and smoking, personal preferences are more important than the results of scientific studies.

Test Homogeneity: KR-20 coefficients for TOSA, CCS, and ICS, and the alpha reliability coefficient for TOLI are given in the following list:

TOSA - 0.55
CCS - 0.45
ICS - 0.39
TOLI - 0.57

The KR-20 coefficients for the two subtests are lower than the KR-20 coefficient for the total test. However, the above coefficients indicate that the two subtests are slightly more homogeneous than the total test. If 20 questions were selected from the total test at random, one might expect the KR-20 coefficient for these 20 questions to drop below 0.39 simply because of the decrease in test length.

The above coefficients indicate that the CCS is the most homogeneous of the three multiple-choice tests and TOLI is more homogeneous than the multiple-choice tests. However, Cronbach (1950, p. 4) indicates that the alpha coefficient for Likert-item tests may be spuriously high. The fact that every question has the same set of response categories may contribute to the measured homogeneity of the test. Response biases as well as item content may be contributing to the measured homogeneity.

The above coefficients are quite low which indicates that more extensive item revision and selection procedures would be desireable. A number of factors which are not related to the content of the items may have contributed to lowering the above KR-20 coefficients. The effect of the test length, the item difficulty, and the homogeneity of the sample of subjects on measured test-homogeneity are discussed by Gulliksen (1961, pp. 76-126).

If all other factors are kept constant, the reliability coefficient will increase as the number of items in the test increases. This trend has been extensively investigated and the Spearman-Brown formula gives a mathematical relationship between test length and the reliability coefficient (Gulliksen, 1961, pp. 62-86).

In the construction of the test items, no attempt was made to ensure that the item difficulties would be near 0.5. The fact that twelve out of thirty-nine items have difficulty levels above 0.75 or below 0.25 will tend to lower the KR-20 coefficient for the test.

The measured homogeneity for a test tends to decrease as the homogeneity of the subjects increases. The sample used in the present study (grade eleven chemistry and physics students) is a relatively homogeneous group. A certain amount of screening is done before students are allowed to take these courses. That is, they must first pass the corresponding courses in grade ten. In addition, students who are not interested in science tend to avoid these courses since chemistry and physics are not compulsory.

Factor Analysis: Table XII contains the factor pattern matrix for a factor analysis of the thirty-nine test items in the Test On Scientific Attitude. The last column in Table XII contains the communalities of the test items. This represents that portion of the item variance which each item has in common with the other items. This is an oblique solution which was obtained by the rotation procedure developed by Harris and Kaiser (1964). The solution reported is an A'A proportional to L solution which Harris and Kaiser (1964, p. 361) recommend for situations in which factor complexity is expected. A is the factor pattern matrix and L is the matrix of intercorrelations of the factors. The rotation was performed on a factor loading matrix obtained by the unweighted least squares factoring of the tetrachoric correlation matrix given in Appendix F. A more detailed discussion of the above procedures is presented in Chapter III. The intercorrelations for these nine factors are given in Appendix G.

The decision to retain nine factors was made on the basis of a scree test (Cattell,1966) and by comparing the nine-factor solution with solutions for seven, eight, and twelve factors. Because of the cost for obtaining solutions with higher numbers of factors, ten- and eleven-factor solutions were not obtained. For the scree test, the eigen roots of the correlation matrix were plotted and the most noticeable break in the curve occurred after the ninth largest root.

Table XII.

OBLIQUE FACTOR PATTERN MATRIX FOR NINE FACTORS OF THE THIRTY-NINE TEST ITEMS IN TOSA

ITEM	FACTOR 1	2	3	4	5	6	7	8	9	COMM.
1	-04	14	14	13	13	12	10	56*	-05	45
2	-01	08	19	-04	29*	06	13	-56	-09	47
3	-07	-07	15	19	33*	-38	02	25*	06	40
4	10	11	06	-12	-26	05	27*	-22	18	26
5	14	-07	12	07	24*	27*	13	06	04	24
6	47*	06	-00	15	29*	10	01	-02	-13	38
7	07	01	04	-04	60*	-02	09	-07	01	40
8	-11	04	03	51*	04	03	-09	16	-02	32
9	05	02	-00	-01	07	06	04	04	59*	38
10	06	07	18	14	47*	11	03	-05	36*	55
11	02	-04	01	-02	-11	11	-18	-01	39*	20
12	08	12	-16	15	25*	01	02	-07	06	16
13	06	-03	05	03	05	10	79*	01	-04	65
14	-02	-08	-05	07	10	-06	14	12	01	07
15	04	-05	-02	58*	-02	-07	-04	10	06	36
16	-04	-09	12	-07	-03	-21	14	09	38*	23
17	16	-20	-08	32*	31*	-11	15	01	15	37
18	10	-02	-04	34*	04	-07	26*	-07	17	27
19	09	-14	05	13	-58	-01	-02	-11	11	38
21	-31	28*	-05	14	11	-04	16	04	01	21
22	61*	29*	-01	-09	-24	-58	03	-03	-48	99
23	01	-06	49*	04	-05	07	01	-06	05	25
24	44*	-06	-20	-04	02	08	-02	-03	06	24
25	04	03	02	-14	09	21*	-06	-12	-00	09
26	74*	11	28	01	-02	11	27*	04	37*	99
27	01	14	08	-05	10	-11	31*	-01	-07	17
28	-03	65*	-05	05	07	-06	08	-07	-09	45
29	-18	31*	67*	-06	03	04	16	05	04	62
30	-09	09	-03	01	14	06	15	-07	21	13
31	07	52*	09	-21	07	-17	-22	-03	24	49
32	32*	69*	-07	-16	13	29*	-06	36*	03	95
33	15	38*	-26	33*	07	-17	-16	-25	72*	99
34	-11	30*	11	06	-09	-00	17	02	25	21
35	18	19	11	26*	09	08	-01	-16	-19	23
36	17	-28	57*	-03	24	-29	-17	-01	-08	60
37	16	11	18	27*	-14	13	-26	-12	-04	24
38	20	02	-02	-41	03	-05	-05	05	32*	30
39	14	-07	02	01	-05	55*	-06	12	04	35
40	-00	16	13	12	01	16	-23	-17	-22	19

* Marks the salient item loadings on each of the factors. The entries in the above matrix have been multiplied by a factor of 100.

When twelve factors were retained, two of these factors had only one large loading and a third factor contained high loadings mainly for variables which had grouped together on one of the factors in the nine-factor solution. The seven-factor solution contained factors with a large number of high loadings and also contained a number of variables which had no high loadings on any factor. The most noticeable difference between the eight-factor solution and the nine-factor solution is factor three in Table XII. This combination of test items is not apparent in the eight-factor solution although most of the other factors are quite similar.

There appears to be some indication that it is probably not meaningful to retain more than nine factors. In addition to the points discussed above, the sum of the nine eigen roots associated with these factors was compared with the sum of the communalities of the test items. The obtained ratio is 0.982 which indicates that these nine factors account for 98.2 percent of the total common variance of the test items. This may be an indication that the variables have been slightly overfactored. That is, fewer than nine factors may adequately represent the test items. However, the nine-factor solution, presented in Table XII appears to be more satisfactory than solutions with seven and eight factors. A number of the factors in the nine-factor solution have a fairly large number of high loadings which would indicate that extensive factor splitting has not occurred.

Following is a list of test items which make the greatest contribution to each factor:

Factor I: 6, 22, 24, 26, 32
Factor II: 21, 22, 28, 29, 31, 32, 33, 34
Factor III: 23, 29, 36
Factor IV: 8, 15, 17, 18, 33, 35, 37
Factor V: 2, 3, 5, 6, 7, 10, 12, 17
Factor VI: 5, 25, 32, 39
Factor VII: 4, 13, 18, 26, 27
Factor VIII: 1, 3, 32
Factor IX: 9, 10, 11, 16, 26, 33, 38

The decision to include or exclude certain test items in the list of salient items for each factor was made on somewhat subjective grounds in some cases. All items with loadings of 0.30 and greater were included. In the consideration of loadings between 0.20 and 0.29, the following points were examined: the difference between the value of the loading being considered and the value of the next highest loading on the factor, the number of items already included in the factor, the size of the other loadings for the item being considered, and the extent to which the item being considered appears to be related to the other items in the factor. The application of these guidelines to the above solution is dicusssed under the detailed discussion of each factor below.

In the following paragraphs, the above solution is first disucssed in general terms and then each factor is discussed in detail.

General Discussion: Items 14, 19, 30 and 40 do not have high loadings on any of the factors. Although item 25 is listed under factor VI, its highest loading is 0.21. All of these items except 19 have very low communalities. That is, they have very little in common with other items. This lack of relationship with the rest of the test is also indicated by low biserial correlations for items 14, 19, 25 and 40 (see the results of the item analysis reported earlier in this Chapter). Therefore, it is not likely that these items would show high loadings if additional factors were obtained.

Question 19 is discussed in the first section of this chapter. It appears to be a trick question and the possible reasons for selecting the keyed response bear little relationship with the behavioral specifications of the attributes. Question 30 asks the student to express an opinion as to whether or not the theory of evolution should be discussed in biology class. This question was designed to measure objectivity; however, a student's respect for the individual rights may influence his response to this question. Examination of questions 14, 25 and 40 does not reveal any apparent reason why these items are not more closely related to other items in the test.

The rationale which served as a guideline for the construction of the test items suggests two possible criteria on which the items might be classified. One classification which has been made is the division of the test items into the cognitive and intent component subtests (questions 1 to 19 are cognitive items; questions 21 to 40 are intent items). A second grouping of the items, according to which attribute they were designed to measure, is given in Chapter III. Because of the fact that the behavioral specifications used to define the attributes are often repeated under two or three attributes, most of the questions are listed under more than one attribute. This list is repeated below:

Critical mindedness (questioning attitude) -
9, 19, 21, 24, 25, 31, 32, 36

Supended judgement (restraint) -
1, 2, 4, 6, 7, 11, 18, 25, 26, 27, 28, 34, 35, 37

Respect for evidence (reliance on fact) -
10, 11, 12, 14, 16, 26, 27, 28, 31, 32, 33, 38

Honesty -
3, 8, 18, 22, 33, 39

Objectivity (open-mindedness) -
8, 13, 15, 17, 23, 29, 30, 36, 40

Willingness to change opinions -
1, 3, 4, 5, 6, 10, 29, 30, 37

All of the questions in factor II and factor III are from the Intent Component Subtest and all of the questions in factor V are from the Cognitive Component Subtest. Of the five questions in factor I, all but one are from ICS. Factor VI is composed of three questions from ICS and one question from CCS. Factor VIII is composed of two questions from CCS and one question from ICS. Factors IV, VII and IX contain approximately equal numbers of questions from both subtests. This distribution of the questions among the factors indicates that the two subtests do not measure the same characteristics and lends considerable support to the original division of the questions into the two subtests.

Since the test questions are often listed under more than one attribute, a certain amount of agreement between this list and the list of items in each factor is almost inevitable. However, the extent of the agreement is sufficient to give some support to the original classification. Each of the nine factors can be associated with one of the six attributes. In the following paragraphs, the list of items in each of the factors is discussed in relation to the classification of the test items on the basis of the definitions of the attributes.

Factor I: Questions 6, 24, and 26 are all listed under suspended judgement. The keyed response for question 32, ask a friend to present facts and arguments to support a statement, is related to critical mindedness and respect for evidence. However, the other distractors are related to suspended judgement. Question 22 does not appear to have a great deal in common with the other questions in this factor. This question, which deals with interpretation of data, is somewhat process oriented. However, alternative C, take more measurements when conflicting results are obtained, is related to suspended judgement and respect for evidence.

Factor II: Questions 21 and 22 with loadings of 0.28 and 0.29 were included in this factor because there are a large number of loadings near 0.30 on this factor. Questions 28, 31, 32 and 33 are listed under respect for evidence. Question 31 (questioning religious beliefs because scientists have cast doubt on some of them) appears to be more related to objectivity, but a respect for evidence is also implied in this response. Question 22 (discussed under factor I) can also be related to respect for evidence. Questions 29 (one should be willing to admit that there may be some truth to certain superstitions) and question 34 (the cause of the common cold is not known for certain) do not appear to be related to respect for evidence. Question 29 is related to questions 21, 28, and 31 in that all these questions refer to controversial topics (superstitions, religion, marijuana, and pollution).

Factor III: All the test items in this factor (29, 33, and 36) are listed under objectivity. These three questions refer to fluoridation, superstitions and astronomy. Question 26 with a loading of 0.28 was not included in this factor because the other three loadings are all considerably higher (0.49 - 0.67) and because question 26 has been included in other factors. Question 26 which refers to inconsistent lab results in the test for starch in leaves does not appear to have very much in common with questions 33, 29, and 36.

Factor IV: Questions 35 and 37 with loadings of 0.26 and 0.27 are included in this factor because a number of the other loadings on this factor are near 0.30. The loadings of 0.26 and 0.27 are the largest loadings for questions 35 and 37, and these questions are not included in any of the other factors.

Questions 15, 18, 35 and 37 are all listed under suspended judgement. In question 8 Schleiden's failure to account for all of his observations, demonstrates dishonesty and a lack of objectivity. However, this may also be interpreted as forming generalizations not justified by available data which demonstrates a lack of suspended judgement. Questions 17 and 33 do not appear to be related to suspended judgement.

Factor V: Since a number of the loadings on this factor are near 0.30, questions 5 and 12 with loadings of 0.24 and 0.25 are included. Question 12 does not have salient loadings on any of the other factors. Question 5 is also included in factor VI, but its loading on factor VI (0.27) is not much higher. Question 36 with a loading of 0.24 was not included because this question has a much higher loading on factor III and it is more related to the other questions in factor III than it is to the questions in factor V.

Questions 3, 5, 6, 10, and 12 which are included in this factor are all listed under willingness to change opinions. Questions 3, 5, 6, and 12 all deal with adjusting theories to account for new data. The philosophers described in question 10 were not willing to accept Galileo's explanations. Question 7 which deals with generalizing beyond the scope of available data is more related to respect for evidence and suspended judgement than to willingness to change opinion. Question 17 (evaluating ideas which disagree with one's hypothesis) is more related to objectivity. Question 2 which describes a scientist reacting to a report on a type of water that boils at 450°F is more related to suspended judgement.

Factor VI: This is a bipolar factor. There are the same number of high negative loadings as there are high positive loadings. This factor was not reflected because all of the high negative loadings are for items which are included in other factors. The loadings for the items which are included in this factor have a fairly wide range (0.21 - 0.55). Question 25 with a loading of 0.21 is included because this loading is still reasonably close to the loadings for questions 5 and 32 and because question 25 does not have high loadings on any of the other factors.

The questions in this factor appear to be most closely related to critical mindedness. The list for critical mindedness includes questions 25 and 32. Priestley's behavior as described in question 5 (he did not accept the new theory of combustion even though there was considerable evidence to support it) demonstrates unwillingness to change opinions and a lack of respect for evidence. However, he also failed to demonstrate critical mindedness in his evaluation of the phlogiston theory. Question 39 (repeating a chemistry experiment after adding a wrong solution) does not appear to be related to critical mindedness.

Factor VII: Questions 4, 18, 26, and 27 in this factor are all listed under suspended judgement. The loadings for all these questions are near 0.30. Question 13 which is also included in this factor has a loading of 0.79 which is considerably higher than the loadings for the other four questions in this factor. The situation described in question 13 (reasons why Arrhenius' theory of ionization was not widely accepted) is more relevant to objectivity than to suspended judgement. However, Arrhenius demonstrated critical mindedness and suspended judgement in his search for a new theory.

Factor VIII: This is a bipolar factor with three large positive loadings and three large negative loadings. This factor was not reflected because question 1 which is not included in any of the other factors has a high positive loading on this factor. The second and third highest positive loadings are slightly higher than the second and third highest negative loadings.

Questions 1 and 3 are both relevant to willingness to change opinions. These two questions both refer to changing theories to explain new evidence. Question 32 is included in three other factors. The keyed response to this question (ask a friend to provide evidence and arguments to support a statement) appears to be more related to critical mindedness and respect for evidence than it is to willingness to change opinions. However, the other alternatives reflect a willingness to consider ideas presented by others and this is one of the defining behaviors of willingness to change opinions.

Factor IX: Since factor IX has a large number of positive loadings near 0.40, test items with loadings in the range 0.20 - 0.25 were not included in this factor. Questions 31 and 34 with loadings of 0.24 and 0.25 respectively are included in other factors. Question 30 with a loading of 0.21 has not been included in any of the other factors.

The questions in this factor reflect respect for evidence. Questions 11, 16, 26, 33 and 38 have all been classified under this attribute. Question 9 is classified under suspended judgement and 10 is classified under willingness to change opinions. However, the philosophers' actions described in these questions demonstrate a lack of respect for evidence. That is, they were not willing to evaluate Galileo's findings.

Of the questions which have loadings in the range of

0.20 - 0.25 on this factor, questions 21 and 34 are included in factor II which is also a factor which reflects respect for evidence. Question 30 on the teaching of the theory of evolution in a biology class, is not relevant to respect for evidence.

Summary: The following list matches each factor with the affective attribute to which it is most closely related:

Factor I - suspended judgement
Factor II - respect for evidence
Factor III - objectivity
Factor IV - suspended judgement
Factor V - willingness to change opinions
Factor VI - critical mindedness
Factor VII - suspended judgement
Factor VIII - willingness to change opinions
Factor IX - respect for evidence

Honesty is not represented in the above list. A possible reason for this is that this attribute was not defined very distinctively. Two of the behavioral objectives defining honesty are very similar to behavioral objectives which define objectivity and critical mindedness. These are the behaviors related to the evaluation and reporting of data which is contradictory to predicted hypotheses. The other behavioral objective defining honesty (acknowledges work done by others) is not easily translated to a suitable test question.

The researcher does not propose that the test items be divided into subtests to represent each of the affective attributes. In fact, the overlap of behavioral objectives used to define the attributes makes it impossible to give a distinct division of the test items. In the factor solution presented above, eleven of the questions are included in more than one factor. Considerable overlap was predicted from the classification of the test items given in Chapter III.

The purpose of the factor analysis, rather than to provide subtests, is to provide an empirical verification of the theoretical classification of the test items. The agreement between the theoretical classification and the classification by the factor solution is reasonably good. A one hundred percent agreement between the two classifications was not expected. In classifying the items on a theoretical basis, it is impossible to anticipate all of the factors which might influence a respondent's decision to select a particular response. The items were classified on the basis of the most obvious relationships with the behavioral objectives used to define the attributes. In an attempt to interpret the factor solution, some of the less obvious relationships were revealed.

Each of the nine factors has been matched with one of the affective attributes (page 98). The list of items included in each of these factors is given on page 89 and the theoretical classification of the test items is given on page 91. Of the fifty entries in the first list and the fifty-eight entries in the second list, there are thirty-one instances of agreement between the two classifications. As is indicated above in the detailed discussion of each factor, another ten of the entries in the factor list can be explained when the questions are closely examined to determine factors which might influence the selection of alternatives other than the keyed response. When each factor is identified with one of the attributes, it is possible to relate approximately 80% of the salient factor loadings to the item classification which was based on the definitions of the attributes.

IV. Test Stability

Test-retest results for the Test On Scientific Attitude were obtained for 105 students who were tested during the fall semester. The second testing was done three weeks after the first test date. These results were used to examine the test stability of TOSA, CCS, and ICS as measured by the test-retest correlation coefficients. The following test-retest correlations were obtained:

TOSA - 0.71
CCS - 0.68
ICS - 0.64

The test-retest correlation coefficients are considerably higher than the KR-20 coeficients reported earlier in this chapter (0.55, 0.45, and 0.39 for TOSA, CCS, and ICS, respectively). Although the KR-20 coefficients are somewhat low, the test-retest correlations of stability are satisfactory for all three tests. The test-retest correlation or estimate of stability is probably a more important criterion for the three tests.

The test-retest, phi correlation coefficients for the test items in TOSA are reported in Table XIII.

Table XIII.

TEST-RETEST CORRELATIONS FOR 39 TEST ITEMS IN TOSA FOR 105 STUDENTS

ITEM	CORRELATION	ITEM	CORRELATION	ITEM	CORRELATION
1	0.41	14	0.34	28	0.66
2	0.51	15	0.03*	29	0.20
3	0.42	16	0.26	30	0.58
4	0.30	17	0.40	31	0.35
5	0.27	18	0.38	32	0.51
6	0.27	19	0.17*	33	0.39
7	0.30	21	0.53	34	0.58
8	0.42	22	0.34	35	0.16*
9	0.37	23	0.26	36	0.41
10	0.29	24	0.41	37	0.30
11	0.28	25	0.52	38	0.40
12	0.30	26	0.71	39	0.58
13	0.31	27	0.12*	40	0.20

* Not significantly different from 0.0 at the 0.05 probability level

The test-retest correlations for the three tests may be increased if the questions which have low test-rest correlations were revised, omitted, or replaced by other questions.

The weaknesses in question 19 are discussed earlier in this chapter under the headings of content validity and item analysis. There are no obvious reasons why questions 15, 27, 35, and 40 should have such low test-retest correlations.

The means and standard deviations for the test-retest data on TOSA, CCS and ICS are summarized in table XIV.

Table XIV.

MEANS AND STANDARD DEVIATIONS FOR TEST-RETEST DATA ON TOSA, CCS, AND ICS FOR 105 STUDENTS

TEST	MEANS		ST. DEV.	
	TEST	RETEST	TEST	RETEST
TOSA	52.8	52.9	11.2	10.9
CCS	52.0	54.0	15.7	14.8
ICS	53.4	52.0	13.4	12.4

There are no apparent trends in the means from test to retest. The standard deviations all decrease slightly from test to retest. All differences, particularly on the total test, are small and it should be safe to assume that the initial testing did not have any meaningful effect on the retest results.

V. External Validity

The external validity of the test items in TOSA is examined in relation to teacher ratings of student behavior (see Appendix E for instructions to the teachers). Teachers were asked to rate their students on a four-point scale on the basis of the extent to which the student exhibited the behaviors which were used to define the attributes. One rating on the overall set of behaviors was obtained for each student. Since teacher ratings were obtained only for those students who wrote during the spring semester, the test results for those students who wrote during the fall semester are not included in the analysis reported in this section.

Description of the High and Low Student Groups: These two groups are defined in Chapter I as the top and bottom twenty percent of the students in each class based on the teacher ratings. Since analysis of covariance with scholastic ability as the covariate is used to test for differences between these two groups, students for whom SCAT scores were not available were not included in these two groups. Out of the sample of 156 students, 30 students were assigned to each group.

Out of the 156 students, thirty-one were given a rating of 4 (frequently exhibits the behaviors listed in Appendix E), fifty-four were given a rating of 3, fifty-three were given a rating of 2, thirteen were given a rating of 1, and five students were not given any rating because the teacher felt that he did not know these students well enough. The high student group consists of twenty-one students who had received a rating of 4 and nine students who had received a rating of 3. The low student group consists of nine students who had received a rating of 1 and twenty-one students who had received a rating of 2. The students in the high and low groups with ratings of 3 and 2 were selected randomly.

Tests for Differences Between Groups: One-way analysis of covariance with scholastic ability as the covariate was used to test the following null hypotheses:

Hypothesis Ia: There is no significant difference between the mean score on TOSA for the high student group and the mean score on TOSA for the low student group when scholastic ability is the covariate.

Hypothesis Ib: The above hypothesis stated for CCS scores.

Hypothesis Ic: The above hypothesis stated for ICS scores.

Hypothesis Id: The above Hypothesis stated for TOLI scores.

The results of these analyses are summarized in tables XV to XVIII.

Table XV.

ANALYSIS OF COVARIANCE ON TOSA SCORES
LOW STUDENT GROUP VS HIGH STUDENT GROUP

SOURCE	DF	MEAN SQUARE	F	P
Between Groups	1	1723	21.5	<.001
Error	57	80.09		

	SAMPLE SIZE	GROUP MEANS		
		UNADJ	ADJ	SCAT
LOW	30	45.7	46.2	71.7
HIGH	30	59.7	59.2	85.2

Tables XVI.

ANALYSIS OF COVARIANCE ON CCS SCORES
LOW STUDENT GROUP VS HIGH STUDENT GROUP

SOURCE	DF	MEAN SQUARE	F	P
Between Groups	1	1442	14.5	<.001
Error	57	99.19		

	SAMPLE SIZE	GROUP MEANS UNADJ	ADJ	SCAT
LOW	30	47.6	49.1	71.7
HIGH	30	62.5	61.0	85.2

Table XVII.

ANALYSIS OF COVARIANCE ON ICS SCORES
LOW STUDENT GROUP VS HIGH STUDENT GROUP

SOURCE	DF	MEAN SQUARE	F	P
Between Groups	1	2055	12.9	0.001
Error	57	159.5		

	SAMPLE SIZE	GROUP MEANS UNADJ	ADJ	SCAT
LOW	30	43.8	43.4	71.7
HIGH	30	57.2	57.6	85.2

Table XVIII.

ANALYSIS OF COVARIANCE ON TOLI SCORES
LOW STUDENT GROUP VS HIGH STUDENT GROUP

SOURCE	DF	MEAN SQUARE	F	P
Between Groups	1	332.5	12.0	0.001
Error	57	27.76		

		GROUP MEANS		
	SAMPLE SIZE	UNADJ	ADJ	SCAT
LOW	30	79.7	80.1	71.7
HIGH	30	86.2	85.8	85.2

In the above analyses, the probabilities that the assumption of homogeneity of regression is satisfied are 0.92, 0.54, 0.54, and 0.81 for TOSA, CCS, ICS, and TOLI respectively.

Conclusions: Hypotheses Ia, Ib, Ic, and Id can be rejected at the 0.001 level of significance. The mean scores for the high student group on TOSA, CCS, ICS, and TOLI are higher than the mean scores for the low student group. The division of the students into the two groups on the basis of the teacher ratings results in two groups of students who are significantly different with respect to their mean scores on TOSA, CCS, ICS, and TOLI. For TOSA, CCS, and ICS the differences between the adjusted means for the two groups are quite large (13, 12, and 14 respectively).

It is difficult to evaluate the relative abilities of the Test On Scientific Attitude and the Test Of Likert Items to discriminate between the high and low student groups. Because of the different scoring systems used (each question in TOSA is scored 1-0 where one alternative is the keyed response while the four alternatives for each statement in TOLI is scored 4-3-2-1), one would expect the mean scores for the high and low student groups on TOLI to be higher and closer together than the mean scores on TOSA. However, from a practical point of view, TOSA would provide a greater opportunity to observe a meaningful score difference between samples if these tests were used to evaluate the relative effectiveness of certain teaching methods and materials in developing the attributes which these tests were designed to measure.

In the overall comparison of the two test formats, consideration must be given to the response biases which are associated with the Likert-scale format (see Chapter II). The researcher feels that the multiple-choice format more adequately meets the requirements specified in the rationale developed in this study. This format more readily lends itself to the description of courses of action relevant to the behavioral objectives used to define the attributes. It is possible to give a more detailed description of classroom situations in a multiple-choice question. A survey of the various attitude scales discussed in Chapter II reveals that a major portion of the statements in these scales are not related to classroom situations.

Comparison of the Two Groups on Each Test Item: A number of statistics on each test item in TOSA relative to the high and low student groups are summarized in Table XIX. These include the item means for each group and the difference between the item means for the two groups. The last column in the table contains the tetrachoric correlations between the test items and the dichotomous grouping of the students in which the students in the high group were assigned a 1 and the students in the low group were assigned a 0. The tetrachoric correlations were calculated by the cosine-pi formula and are therefore biased in the case of test items which have extreme difficulty levels (Ferguson, 1959, p. 244). The difficulty levels for each item are also listed in the table.

Tests for significant differences have not been made; however, some general statements concerning the relative abilities of the test items to discriminate between the two groups can be made. The high group does not have a higher mean than the low group on questions 4, 15, 16, 19, 22, 23, 25 and 40. The differences between the means for the high group and the means for the low group are quite large for questions 17, 18, and 28. The differences for items 5, 6, 7, 8, 12, 13, 24, 26, 29, 30, and 35 are all higher than or equal to 20 and are probably large enough to be considered meaningful.

Table XIX.

HIGH-LOW GROUP COMPARISONS FOR EACH TEST ITEM IN TOSA

ITEM	DIFFICULTY	ITEM MEAN		DIFFERENCE	CORRELATION
		HIGH	LOW		
1	85	93	83	10	39
2	61	67	60	07	11
3	59	73	60	13	23
4	52	47	47	00	00
5	66	80	57	37	41
6	55	77	53	24	39
7	52	73	47	36	42
8	32	50	20	30	50
9	82	90	80	10	31
10	80	83	67	16	35
11	54	57	43	14	21
12	55	73	50	23	38
13	22	40	13	27	52
14	40	40	37	03	05
15	26	27	33	-06	-12
16	50	47	50	-03	-05
17	58	90	47	43	73
18	42	67	27	40	59
19	18	13	30	-17	-38
21	20	27	10	17	44
22	11	10	10	00	00
23	37	27	33	-06	-12
24	64	77	47	30	48
25	35	43	43	00	00
26	87	97	77	20	70
27	58	57	40	17	26
28	68	90	53	47	68
29	60	60	40	20	31
30	49	63	43	20	31
31	81	87	70	17	38
32	86	87	67	20	43
33	85	97	80	17	66
34	39	43	27	16	28
35	50	63	37	26	41
36	18	17	10	07	23
37	39	47	40	07	11
38	56	67	53	14	22
39	56	53	50	03	05
40	34	33	47	-14	-22

Although the means for the high group in this sample are higher than the means for the low group on questions 2, 14, 36, 37, and 39, the differences for these questions are quite low and are possibly not significantly different from zero.

From the information presented in Table XIX, it would appear that a number of the individual test items do not effectively discriminate between the high and low student groups.

Correlations of Test Scores with Teacher Ratings: Groups of 62, 51, and 38 students were assigned ratings by three teachers (each group was rated by one teacher). The three groups are not equivalent with respect to TOSA and SCAT scores and the mean teacher ratings for the three groups are not equivalent. Since it is possible that the three teachers may have associated different meanings with the categories which were used to assign ratings (see Appendix E), correlations were calculated separately for the three groups. Correlations with teacher ratings were calculated for TOSA, CCS, ICS, TOLI, SCAT, STEP, and the final science marks (FSCM) which were assigned by the teachers at the end of the school term. The correlations are given in Table XX. Average correlations for the three groups were calculated by applying the Fisher-Z transformation to the correlations, taking the average of the three transformations, and converting this average back to a correlation (Ferguson, 1959, p. 412).

TABLE XX.

CORRELATIONS OF TEST SCORES WITH TEACHER RATINGS

TEST	GROUP SIZE			AVERAGE
	62	51	38	
TOSA	0.43*	0.20	0.08	0.25
CCS	0.26#	0.10	0.04	0.16
ICS	0.39*	0.23	0.09	0.25
TOLI	0.20	0.37*	0.27	0.28
SCAT	0.52*	0.27	0.06	0.30
STEP	0.38*	0.26	0.06	0.24
FSCM	0.41*	0.67*	0.33#	0.49

* Significantly different from 0.0 at the 0.01 probability level

# Significantly different from 0.0 at the 0.05 probability level

The most consistent correlation across the three groups is the correlation of teacher rating with final science mark, and the average correlation for final science mark is considerably higher than the average correlations for the other scores. considerably higher than the other correlations. It appears that the teachers used student achievement in science as a major criterion when assigning ratings.

Since the questions in ICS are worded in terms of a student's activities and the questions in CCS are worded in terms of a scientist's activities, one might expect scores on ICS to be more highly correlated with the teacher ratings of student behavior. This is supported by the correlations in Table XX.

There is no consistent trend in the differences between the correlations for TOLI and TOSA. Although there are considerable differences between these correlations for each of the three groups, these differences are not consistent from one group to another. Consequently, there is little difference in the average correlations for TOLI and TOSA with teacher ratings.

Teacher rating of student behavior is probably not an appropriate criterion to use to examine the validity of TOSA. The correlations in Table XX indicate that the validity of the teacher ratings may be somewhat questionable. In addition, it is possible that the cognitive and intent components of attitudes may not be highly correlated with the action component. A student's behavior in a given situation may vary considerably from his expressed intentions with respect to that or some similar situation.

The information provided in the present study does not account for the wide variation in correlations for the three teachers. It is possible that the teachers may have interpretted the instructions (Appendix E) differently. It is also possible that the three teachers may not have been equally aware of the affective development of their students. Since the group sizes are quite small, some of this variation will be due to sampling error.

Chapter V

Summary And Recommendations

I. Summary

The procedures followed in the present study were designed to answer the major question posed in Chapter I. That is, is the rationale outlined in Chapter III a practical approach to evaluation in the affective domain of science education. These procedures included the definition of a set of affective objectives, the construction of test questions related to these objectives, and the field-testing of these questions.

The Nay-Crocker inventory of affective attributes of scientists as defined in terms of student behaviors was used as a framework for affective objectives in science education. A selected set of attributes (critical mindedness, suspended judgement, respect for evidence, honesty, objectivity, and willingness to change opinions) was defined in terms of student behaviors and test questions were constructed to reflect the defining behaviors.

Cognitive, intent, and action components of the affective attributes were defined. Questions were constructed to measure the cognitive and intent components (Test On Scientific Attitude - TOSA), and teacher ratings of student behaviors were obtained to provide information relevant to the action component. The questions in the Cognitive Component Subtest (CCS) were designed to measure the student's understanding of how the behaviors which define the attributes are manifest in the activities in which scientists participate. The questions in the Intent Component Subtest (ICS) were designed to measure the student's behavioral intent with respect to the attributes. These questions require the student to indicate a preference for a given course of action in a given situation. The format of the questions is consistent with the behavioral specification of the attributes and with the definition of attitude as a predisposition to some preferred response.

Each test question in TOSA was designed to reflect selected behavioral definitions so that it was possible to relate each question to one or more of the six attributes. In this classification of the questions (page 91), questions were frequently listed under more than one attribute because some of the defining behaviors were listed under more than one attribute and some questions were related to more than one defining behavior.

The Test Of Likert Items (TOLI) was constructed to provide a comparison between this test format and the multiple-choice format described above. The correlations of TOLI with TOSA is 0.37. Since only those opinion statements which were considered to contain content which was similar to the content of TOSA were included in TOLI, this correlation is quite low. Although the content of the statements in TOLI is quite similar to the content of the questions in TOSA, the context is not. In TOLI, each statement is considered separately from the others while each alternative in the questions in TOSA must be considered in relation to three other distractors and in relation to the situation described in the stem. The format of the test questions in TOSA is more consistent with the rationale outlined in Chapter III than is the format of the statements in TOLI. Since the stem of a multiple-choice question can be more extensive than a statement in a Likert scale, it is possible to give a more detailed description of the situation in a multiple-choice question.

Item analysis was used to examine the properties of the individual test questions. The range of the item-difficulty levels is 0.11 to 0.87. Out of the 156 alternatives for the thirty-nine questions, seven received less than three percent of the responses. The biserial correlation coefficients for questions 19 and 22 are essentially zero. The biserial correlation coefficients for nine questions are in the range 0.15 to 0.24. The remaining coefficients are all higher than 0.24. The KR-20 coefficients for TOSA, CCS, and ICS are 0.55, 0.45, and 0.39, respectively. The alpha reliability coefficient for TOLI is 0.57. The test-retest correlations for TOSA, CCS, and ICS are 0.71, 0.68, and 0.64, respectively. Test-retest data were not obtained for TOLI. Although the KR-20 coefficients are quite low, the test-retest correlations for TOSA, CCS, and ICS are satisfactory.

The intercorrelations among the test questions were examined by factor analysis and the solution for nine factors was reported. This factor solution was discussed in relation to the classification of the test questions which was based on the definitions of the six attributes. When each of the factors was identified with one of the attributes, it was possible to relate approximately 80% of the salient factor loadings to the item classification which was based on the definitions of the attributes. Suspended judgement was identified with three of the nine factors, willingness to change opinion and respect for evidence were each identified with two factors, and objectivity and critical mindedness were each identified with one factor.

The factor solution gave some support to the division of the questions into the cognitive and intent subtests. Two of the factors contain only questions from ICS, and one factor contains only questions from CCS. One factor contains four questions from CCS and one question from ICS, one factor contains three questions from ICS and one question from CCS, and one factor contains two questions from CCS and one question from ICS. The remaining three factors contain approximately equal numbers of questions from both subtests.

The relationships between TOSA and a number of other tests were examined on the basis of the intercorrelations among the following tests: TOSA, CCS, ICS, TOLI, WCTA (critical thinking ability), SCAT (general scholastic ability), and STEP reading. The correlation coefficient between ICS and CCS is 0.23 for 307 grade eleven science students. This low correlation lends further support to the division of the questions into the two subtests. The correlation coefficients for TOSA, CCS, ICS, and TOLI with SCAT are 0.33, 0.44, 0.09, and 0.30, respectively. The correlation coefficients for these four tests with STEP reading are 0.35, 0.41, 0.13, and 0.27, respectively. The correlation coefficients for these four tests with WCTA are 0.41, 0.45, 0.25, and 0.36, respectively. On the basis of the content of the questions in the two subtests (many of the questions in ICS require the expression of personal preference or opinion), it is reasonable to expect CCS to be more highly correlated with the general abilities measured by SCAT, STEP, and WCTA. The correlations of TOSA, CCS, and ICS with TOLI are 0.37, 0.37, and 0.24, respectively. The fact that TOLI has a lower correlation with ICS than with CCS indicates that TOLI may have a fairly strong cognitive component.

Information relevant to the action component of the affective attributes was obtained from teacher ratings of student behavior. These ratings were used to examine the external validity of the tests which were constructed for this study, and to examine the relationships between the action component and the two subtests of TOSA.

The students were divided into high and low student groups on the basis of these ratings and one-way analysis of covariance was used to test for significant differences between the two groups. Scholastic ability as measured by SCAT was employed as the covariate, and TOSA, CCS, ICS, and TOLI were the separate criterion measures. The two groups were shown to be significantly different ($p \leq 0.001$) for all four criterion measures.

Ratings of behavior were provided for three groups of students by three different teachers (only one teacher rated the students in each group). Correlations between teacher ratings and a number of test scores were calculated for each group (page 111). Except for correlations with final science marks, these correlations were inconsistent across the three groups. The average correlations of teacher ratings with TOSA, CCS, ICS, SCAT, STEP, and final science mark are 0.25, 0.16, 0.25, 0.28, 0.30, 0.24, and 0.49, respectively. It appears that the teachers used student achievement in science class as a major criterion when they assigned ratings.

The rationale outlined in Chapter III provided valuable guidance for the construction of the test questions in TOSA. The points made in the above summary indicate that useful tests can be constructed through the application of this rationale. Although there are discrepancies between the empirical structure which was identified by the factor analysis and the logical structure which was based on the rationale, it is not practical to propose an empirically based rationale on the basis of this study alone. These disrepancies may be due to inconsistencies in the rationale, but they may also be due to the fact that some of the questions in TOSA require considerable revision. Sampling error may also account for some these discrepancies. Therefore, any attempt to establish an empirically based rationale should be based on a large number of studies so that consistent trends can be identified.

II. Implications for Classroom Evaluation

The behavioral specification of affective characteristics which are provided in this study may help teachers to become more aware of the affective development of thier students. The tests can be used to obtain measures of achievement of affective objectives. If these materials prove to be helpful, teachers may be able to extend the procedures outlined in this study to a wider range of affective objectives.

The present study indicates that the cognitive and intent subtests are not measuring the same characteristics. Student understanding of how scientists demonstrate the affective attributes in their work probably is not sufficient to ensure that students will demonstrate these characteristics in their own science work or in everyday situations. Teachers must consider this difference when planning classroom activities and evaluating student development. Teachers must also make some attempt to determine the extent to which student behavior is related to responses on the type of test questions in TOSA.

III. Recommendations

1. Since the statistics reported in Chapter IV indicate that some of the questions in TOSA are unsatisfactory, these questions should be revised or replaced by other questions.

The researcher has identified precautions which should be taken in the application of the rationale to test construction. Attempts at test-item construction should not be made until appropriate behavioral objectives are written to define each attribute. The behaviors which proved to be the most useful guidelines for test construction were those behaviors which one could reasonably expect to observe in practical situtions.

It is recommended that the rationale be used to provide guidance for the construction of test items, and that the data analysis described in this study be used to provide information for item revision and selection. This approach is consistent with Loevinger's discussion of the substantive component of construct validity (1967, p. 97). She suggests that test items should be constructed and included in the initial draft of a test on the basis of logical relevance while the final selection of test items should be made on the basis of empirical findings.

2. Alternate procedures for obtaining information relevant to the action component may be preferable to the one used in this study. One alternative would be observation of student activity by the researcher. However, observation of normal classroom activities may not be practical because of the large amount of this type observation required. It may be possible to construct situations designed to provide an opportunity for students to demonstrate the behaviors which define the attributes being measured.

If tearcher ratings are obtained, it may be profitable to obtain a separate rating for each attribute being measured. The categories which are to be used by the teachers should probably be more specifically defined than the descriptions used in the present study.

It may be possible to use teacher ratings in research designed to examine factors which can help account for variations in correlations between teacher ratings and written test scores. This type of research should provide information to help the classroom teacher become more aware of the affective development of the students and to help the teacher to make some intuitive evaluation of this development.

3. The present study has been confined to a small subset of the affective objectives of concern in science education. All of the attributes on which the study has concentrated are listed under attitudes in the Nay-Crocker inventory. Further research is required to determine whether or not the approach described in this study can be applied to the measurement of interests, operational adjustments, appreciations, values, and other attitudes.

4. Other scoring systems for the questions in ICS may prove to be more appropriate than the 1-0 system used in the present study. Since these questions require the student to express a personal preference, it may be appropriate to score each alternative on a scale which decreases in value from the response which is most consistent with the attribute being measured to the response which is least consistent with this attribute. If this approach is to be employed, a common scale could not be used for all questions. Examination of the questions in this subtest readily indicates that the distance separating the most consistent response and the least consistent response is not the same for all questions.

It may be possible to use Thurstone's scaling procedures (Chapter II) for the set of alternatives for each question. Further research is required to establish whether or not the validity and reliability of the test can be improved by the use of a different scoring system. If the validity and reliability of these tests are improved, a decision would have to be made as to whether or not these gains justify the amount of work required to establish the required scales.

5. The model in Figure I on page 37 should be empirically examined to determine whether or not the behavioral specification of the affective objectives can provide insights into instructional materials and methods which might be employed to foster the development of interest, adjustments, attitudes, appreciations, and values. This will involve research to determine what types of instructional materials and methods can be employed to achieve the objectives. Effective research in this area has been difficult to accomplish because of a lack of appropriate evaluation instruments.

From the above discussion it should be evident that the present study has been merely a first step toward the research required in the area of instructional methods and evaluation in the affective domain. The primary purpose of this study has been to demonstrate a practical approach to evaluation in this area.

BIBLIOGRAPHY

Alberta Department of Education (office of the science coordinator). *Objectives of Secondary School Science for Alberta*, Spring, 1970.

Allen, Hugh, Jr. *Attitudes Of Certain High School Seniors Toward Science and Scientific Careers*. New York: Marstin Press, Inc., 1959.

Barber, B. *Science And The Social Order*. New York: Collier Books, 1962.

Bloom, B. S. (ed.). *Taxonomy Of Educational Objectives, Handbook I: Cognitive Domain*. Ann Arbor: David McKay Co., Inc., 1956.

Bloom, B.S., J. T. Hastings, and G. F. Madaus. *Handbook On Formative And Summative Evaluation Of Student Learning*. New York: McGraw-Hill Book Co., 1971.

Buros, O. K. (ed). *The Seventh Mental Measurements Yearbook*. Highland Park: The Gryphon Press, 1972.

Cattell, R. B. "The Scree Test For The Number Of Factors," *Multivariate Behavioral Research*. 1:245-76, 1966.

Cooperative Test Division of the Educational Testing Service. *Technical Report For The Sequential Tests Of Educational Progress*. Princeton: 1957.

__________. *Technical Report For The Cooperative School And College Ability Tests*. Princeton: 1958.

__________. *1962 SCAT-STEP Supplement*. Princeton: 1962.

Cronbach, L. J. "Response Sets And Test Validity," *Educational And Psychological Measurement*, 6: 475-94, Winter, 1946.

__________. "Further Evidence On Response Sets And Test Designs," *Educational And Psychological Measurement*, 10: 3-31, 1950.

Cronbach, L. J. and P. E. Meehl. "Construct Validity in Psychological Tests," *Psychological Bulletin*, 52: 281-302, 1955.

Diederich, P. B. "Components Of The Scientific Attitude," *The Science Teacher*, 34: 23-4, February, 1967.

Ebel, R. L. *Measuring Educational Achievement*. Englewood Cliffs: Prentice-Hall, Inc., 1965.

Eisner, E. W. "Instructional And Expressive Educational Objectives: Their Formulation And Use In

Curriculum," in *AERA Monograph In Curriculum Evaluation, No. 3*. (R. E. Stake, ed.). Chicago: Rand McNally and Co., 1969. pp. 1-18.

Eiss, A. F. "The NSTA Conferences On Scientific Literacy," *The Science Teacher*, 35:30-2, May,1968.

Eiss, A. F. and M. B. Harbeck. *Behavioral Objectives in The Affective Domain*. Washington: National Science Supervisors Association, 1969.

Ferguson, G. A. *Stastical Analysis In Psychology ~~And~~ Education*. New York: McGraw-Hill Co., 1959.

Fishbein, M. "Attitudes And The Predicttion of Behavior," in *Attitude Theory And Measurement*. (M. Fishbein, ed.). New York: John Wiley and Sons, Inc., 1967. pp. 477-91.

Gulliksen, H. *Theory Of Mental Tests*. New York: John Wiley and Sons, Inc., 1950.

Haberman, M. "Behavioral Objectives: Bandwagon or Breakthrough," *Journal of Teacher Education*, 19: 91-4, Spring, 1968.

Hakstian, A. R. "A Comparative Evaluation Of Several Prominent Methods Of Oblique Factor Transformation," *Psychometrika*, 36: 175-193, June, 1971.

Hakstian, A. R. and K. S. Bay. *User's Manual To Accompany The Alberta General Factor Analysis Program*. Division of Educational Research Services, University of Alberta, 1973.

Harman, H. H. *Modern Factor Analysis*. Chicago: University of Chicago Press, 1960.

Harris, C. W. and H. F. Kaiser. "Oblique Factor Analytic Solutions By Orthogonal Transformations," *Psychometrika*, 29:347-362, 1964.

Hedges, W. D. *Testing And Evaluation For The Sciences In The Secondary School*. Belmont: Wadsworth Pub. Co., Inc., 1966.

Kahn, Paul. "An Experimental Study To Determine The Effect Of A Selected Procedure For Teaching The Scientific Attitude To Seventh And Eighth Grade Boys Through The Use Of Current Events In Science," *Science Education*, 46:115-27, March, 1962.

Kimball, M. E. "Understanding The Nature Of Science: A

Comparison Of Scientists And Science Teachers," Journal Of Research In Science Teaching, 5: 110-20, March, 1968.

Klopfer, L. E. History Of Science Cases. Chicago: Science Research Associates, Inc., 1964.

Klopfer, L. E. and W. W. Cooley, Manual For Administering, Scoring, and Interpreting Scores on TOUS; Form W. Princeton: Educational Testing Service, 1961.

Krathwohl, D. R., B. S. Bloom and B. B. Masia. Taxonomy Of Educational Objectives, Handbook II: Affective Domain. New York: David McKay Co., 1964.

Lauridsen, L. I. and W. S. LaShier. "An Investigation Of The Effects Of ISCS Level One On Selected Affective Variables Of Students." A paper presented at the forty-sixth annual meeting of the National Association for Research in Science Teaching held in Detroit. March 27-29, 1973.

Likert, R. "The Method Of Constructing An Attitude Scale," Archives Of Psychology, 1932, pp. 44-53. in Attitude Theory And Measurement. (M. Fishbein, ed.). New York: John Wiley and Sons, Inc., 1967. pp. 90-5.

Loevinger, J. "Objective Tests As Instruments Of Psychological Theory," in Problems In Human Assessment. (D. N. Jackson and S. Messick, eds.), New York: McGraw-Hill, Inc., 1967

Lowery, L. F. "Development Of An Attitude Measuring Instrument For Science Education," School Science And Math, 66: 494-502, May, 1966.

MacDonald, J. B. and B. J. Wolfson. "A Case Against Behavioral Objectives," Elementary School Journal, 71: 119-28, December, 1970.

Mager, R. F. Preparing Instructional Objectives. Palo Alto: Fearon Publishers, 1962.

Magnusson, D. Test Theory. Palo Alto: Addison-Wesley Publishing Company, 1966.

McAsham, H. H. Writing Behavioral Objectives: A New Approach. New York: Harper and Row Publishers, 1970.

McNemar, Q. Psychological Statistics. New York: John Wiley and Sons, Inc., 1949.

Moore, R. W. and F. X. Sutman, "The Development, Field Test And Validation Of An Inventory Of Scientific Attitudes," Journal Of Research In Science Teaching, 7: 85-94, 1970.

Nay, M. A. and R. K. Crocker, "Science Teaching And The Affective Attributes Of Scientists," Science Education, 54: 59-67, January-March, 1970.

Noll, V. H. "The Habit Of Scientific Thinking," Teachers College Record, 35: 1-9, October, 1933.

__________. Measuring The Scientific Attitude," The Journal Of Abnormal And Social Psychology, 30: 145-54, July-September, 1935.

Obourn, E. S. and L. K. Johnson. "An Analysis And Checklist On The Problem Solving Objective," U. S. Department of Health, Education, and Welfare. Circular No. 381 (rev.), August, 1960.

Oppenheim, A. N. Questionnaire Design And Attitude Measurement. New York: Basic Books, Inc., 1956.

Osgood, C. C., G. E. Suci and P. H. Tannenbaum. The Measurement Of Meaning. Urbana: University Of Illinois Press, 1957.

Plowman, P. D. Behavioral Objectives: Teacher Success Through Student Performance. Chicago: Science Research Associates, Inc., 1971.

Popham, W. J. "Objectives And Instruction" in AERA Monograph ON Curriculum Evaluation, No. 3. (R. E. Stake, ed.). Chicago:Rand McNally and Co., 1969. pp. 32-52.

Rokeach, M. Beliefs, Attitudes and Values. San Francisco: Jassey-Bass Inc., 1968.

Schwirian, P. M. "On Measuring Attitudes Toward Science," Science Education, 52: 172-9, March, 1968.

Scriven, M. "Student Values As Educational Objectives" in Proceedings Of The 1965 Invitational Conference On Testing Problems. Princeton: Educational Testing Service, 1966, pp. 33-49.

Shaw, M. E. and J. M. Wright. Scales For The Measurement Of Attitudes. New York: McGraw-Hill Book Company, 1967.

Thurstone, L. L. "Attitudes Can Be Measured," American Journal Of Sociology, 33:529-24, 1928. in Attitude

Theory And Measurement. (M. Fishbein, ed.) New York: John Wiley and Sons, Inc., 1967.

___________. "The Measurement Of Social Attitude," The Journal Of Abnormal And Social Psychology, 26:249-99, 1931. in Attitude Theory And Measurement. (M. Fishbein, ed.). New York: John Wiley and Sons, Inc., 1967.

Torgerson, W. S. Theory And Methods Of Scaling. New York: John Wiley and Sons Inc., 1958.

Winer, B. J. Statistical Principles In Experimental Design. New York: McGraw-Hill Book Company, 1962.

APPENDICES

APPENDIX A

Affective Attributes of Scientists
(Nay and Crocker, 1970, pp. 61-62)

1. Interests

(The motivation for a person to become a scientist and continue to be one.)

- 1.1 Understanding natural phenomena
  - 1.11 Curiosity
  - 1.12 Fascination
- 1.2 Contributing to knowledge and human welfare
  - 1.21 Excitement
  - 1.22 Enthusiasm
  - 1.23 Ambition
  - 1.24 Pride
  - 1.25 Satisfaction

2. Operational Adjustments

(Primary behaviours which underlie competence and success in science, and performance at recognized standards.)

- 2.1 Dedication or commitment
  - 2.11 Perseverance (persistence)
  - 2.12 Patience
  - 2.13 Self-discipline
  - 2.14 Selflessness
  - 2.15 Responsibility
  - 2.16 Dependability
- 2.2 Experimental requirements
  - 2.21 Systematism (methodicalness)
  - 2.22 Thoroughness
  - 2.23 Precision
  - 2.24 Sensitivity
  - 2.25 Alertness for the unexpected
- 2.3 Initiative and resourcefulness
  - 2.31 Pragmatism (common-sensical)
  - 2.32 Courage (daring, venturesomeness)
  - 2.33 Self-direction (independence)
  - 2.34 Self-reliance
  - 2.35 Confidence
  - 2.36 Flexibility

- 2.37 Aggressiveness

2.4 Relations with peers

- 2.41 Cooperation
- 2.42 Altruism
- 2.43 Compromise
- 2.44 Modesty (humility)
- 2.45 Tolerance

3. Attitudes or Intellectual Adjustments

(Intellectual behaviors which are foundational to the scientist's contribution to or acceptance of new scientific knowledge.)

3.1 Scientific Integrity

- 3.11 Objectivity
- 3.12 Open-mindedness
- 3.13 Honesty
- 3.14 Suspended judgment (restraint)
- 3.15 Respect for evidence (reliance on fact)
- 3.16 Willingness to change opinions
- 3.17 Idea sharing

3.2 Critical requirements

- 3.21 Critical mindedness
- 3.22 Skepticism
- 3.23 Questioning attitude
- 3.24 Disciplined thinking
- 3.25 Anti-authoritarianism
- 3.26 Self-criticism

4. Appreciations

(Relative to the foundations and dynamics of science.)

4.1 The history of science

- 4.11 The social basis or the development of modern science
- 4.12 The "two cultures"
- 4.13 Contributions made by individual scientists
- 4.14 The contritubion made by science to social progress and melioration
- 4.15 The relationship between science and technology
- 4.16 The exponential growth of science

4.2 The nature of science

- 4.21 The process of scientific inquiry

- 4.22 The tentative and revisionary character of scientific knowledge
- 4.23 The strengths and limitations of science
- 4.24 The value of one's own contribution and the debt owing other scientists
- 4.25 The communality of scientific ideas
- 4.26 The esthetics and parsimony in scientific theory
- 4.27 The power of individual and cooperative effort
- 4.28 The power of logical reasoning (rationality)
- 4.29 The causal, relativistic, and probabilistic nature of phenomena

5. Values and/or Beliefs

(In the realm of philosophy, ethics, politics, etc.)

5.1 Philosophy

- 5.11 The universe is "real"
- 5.12 The universe is comprehensible (knowable) through observation and rational thought
- 5.13 The universe is not capricious

5.2 Ethical

- 5.21 Science is amoral but scientists have the responsibility to interpret the consequences of their work
- 5.22 Humanism is the highest ideal

5.3 Social

- 5.31 Science must serve the needs of society
- 5.32 Science flourishes best in a free and democratic society

APPENDIX B

TEST ON SCIENTIFIC ATTITUDE

M. J. Kozlow
University of Alberta

M. A. Nay
University of Alberta

Directions:

1. Each question or incomplete statement is followed by four possible answers. Read each question and decide which is the ONE best answer. Mark your answers on the separate answer sheet. Make certain that the number on the answer sheet corresponds to the number of the question that you are answering.

2. Since each question has only four alternatives, ignore column E on the answer sheet.

3. Do not write in this test booklet.

4. Read each question carefully but do not spend too much time on any one question. Answer all questions.

5. Mark only ONE answer for each question.

Example:

	Answer Sheet
200. A person who dedicates his life to the study of chemistry is a	200. A1 B2 C3 D4 E5

A. Biologist
B. Physicist
C. Chemist
D. Zoologist

The test questions are presented in this Appendix under the headings of Cognitive Component Subtest and Intent Component Subtest. When the test was administered to students, all 40 items were given as one test and the cognitive items were mixed with the intent items.

COGNITIVE COMPONENT SUBTEST

1. Scientists recognize that a scientific theory

2[1] A. should not be changed when it is based on a large amount of data.
12 B. may have to be changed to keep up with a rapidly changing world.
85 C. may have to be changed when new observations are made.
1 D. should not be changed when it explains what happens to nature.

2. A science magazine reports that a scientist produced a type of water that boils at 450°F under one atmosphere of pressure. Another scientist reading this report would probably

4 A. believe the report if it was written by a highly respected scientist.
9 B. disbelieve the report because he would know that water boils at 212°F under one atmosphere of pressure.
27 C. do experiments to try to prove that it is wrong.
61 D. neither believe nor disbelieve the report until other scientists study this problem.

3. When observations are made that do not fit an accepted scientific theory, scientists usually

3 A. try to adjust the observations so that they fit into the theory.
8 B. keep the theory as it is since the new observations cannot be used to improve it.
59 C. try to change the theory so that these observations can be explained.
30 D. discard this theory and develop a new one to explain these observations.

---

[1] The number to the left of each alternative is the percentage of students out of 307 who chose this alternative. The keyed response for each question is underlined.

4. When Einstein published his theory of relativity, another famous scientist was reported to have said, "Dr. Einstein's new theory has shattered many of my scientific beliefs to smithereens." This statement indicates that the scientist

52 A. recognized that scientific knowledge is subject to change.
19 B. held some wrong scientific beliefs without knowing it.
3 C. did not believe in the old theory very strongly.
25 D. did not have sufficient evidence to support his original beliefs.

Questions 5 and 6 refer to the following paragraph.

Priestley and Lavoisier are often referred to as the "fathers of modern chemistry". Both of them accepted the phlogiston theory of combustion (all materials give off a substance called phlogiston when they burn). However, Lavoisier did many experiments on burning and developed our modern theory of combustion in which he said that oxygen is always involved. Priestley never accepted this theory.

5. Which one of the following is generally true about scientists, but was NOT demonstrated by Priestley in the above situation?

8 A. Some scientists believe more strongly in their theories.
15 B. Some scientists go overboard in demanding experimental evidence before changing their ideas.
11 C. Scientists do not have to believe in new theories.
66 D. Scientists accept new theories when they are consistent with experimental data.

6. Which one of the following is NOT true about Lavoisier in the above situation?

55 A. He believed that his theory of combustion would not be changed.
10 B. He recognized that theories are likely to change.
9 C. He was prepared to consider ideas presented by others.
27 D. He developed a new theory to explain new evidence.

Questions 7 and 8 refer to the following paragraph.

The German scientist, Schleiden, published a report on the origin of plant cells (1838). He made several observations on the productive cells of some plants and made the following statements:

> It is an absolute law that every cell takes its origin as a very small vesicle [small bladder] and grows only slowly to its defined size. The process of cell formation which I have just described . . . is that process which I was able to follow in most of the plants which I have studied. Yet many modifications of this development can be observed . . . . Nevertheless, the general law remains incontestable [cannot be questioned] . . .

7. Which one of the following is generally true about scientists but was NOT demonstrated by Schleiden in the above situation?

52 A. Scientists try to avoid making general statements based on limited data.
17 B. Scientists are usually careful to report exactly what they observe.
19 C. Scientists collect large amounts of data in order to develop laws of nature.
12 D. Scientists often ignore observations if they do not quite fit into their theories.

8. Some aspects of Schleiden's theory were later shown to be inaccurate. The most probable reason why his theory was NOT completely accurate is that he

16 A. was not able to obtain modern instruments to use in his investigation.
32 B. did not make his theory explain all of his observations.
31 C. tried to develop a theory to explain the origin of all cells.
21 D. felt that his theory could not be questioned.

Questions 9 and 10 refer to the following paragraph:

Galileo gathered much evidence on stars, motion of objects, etc. which gave rise to ideas contrary to those held by the philosophers of his time. The philosophers forced Galileo to recant some of these ideas (say he was wrong) and stopped him from practicing science.

9. Which one of the following best applies to this situation:

9 A. Galileo should have collected more evidence before

disagreeing with the philosophers.

5 B. Galileo's ideas became wrong when he recanted.

3 C. Galileo should have avoided those investigations which led to disagreement with the philosophers.

82 D. Galileo was justified in questioning the beliefs of the philosophers.

10. In their treatment of Galileo, the philosophers

4 A. showed that they did not have a proper respect for evidence.

4 B. seemed to think that they knew all that there was to know.

12 C. were not willing to change their ideas in the face of new evidence.

80 D. showed all of the above characteristics.

11. Drs. Brown, Jones, and Smith are medical researchers. Each one independently investigated the cancer-producing effect of compounds in tar on rats. Dr. Brown reported that there was no effect. Some time later, both Drs. Jones and Smith reported that these compounds were highly cancer-producing. Which one of the following was probably the MOST important reason for Dr. Brown's results?

7 A. He did not consider all the evidence.

54 B. He did not do a sufficient number of controlled experiments.

25 C. He was in a hurry to report his results first.

14 D. He did not analyze his data properly.

12. If a scientist had to choose between two theories, he would probably support the theory which

2 A. most other modern scientists feel is more likely to be correct.

3 B. has more practical value.

40 C. is based on a larger number of observations.

55 D. explains the available observations more satisfactorily.

13. When Arrhenius first proposed his theory of ionization (salts break up into ions when they dissolve in water), very few scientists were willing to support it. Which one of the following is the MOST probable reasons for this disagreement.

22 A. Arrhenius gave a different interpretation to the observations related to this problem.

5 B. The scientists who would not support this theory were not as imaginative as Arrhenius.

47 C. Arrhenius did not have enough evidence to support his theory.

25 D. The scientists who would not support this theory were less willing to risk criticism.

14. A Scientist was studying an ore from the moon in an attempt to obtain a new metal from it. He made several tests but he did not find evidence of a new metal. However, he did identify a peculiar gas which he obtained during one of the tests. He probably would have

5 A. reported that the ore did not contain a new metal.
40 B. reported that portion of his investigation related to the gas.
4 C. not made any report because he did not solve his problem.
50 D. not made any report until he was able to get another scientist to confirm his identification of the gas.

15. Quite often it is possible to give several different explanations for a particular set of observations. Which one of the following would NOT be generally true about such explanations?

26 A. Only one of these explanations could be the true scientific explanation.
18 B. All other things being equal, the explanation which is the most widely known is likely to be the accepted one.
18 C. The explanation which suggests the greatest possibility for further study is likely to be the one which most scientists use.
38 D. All these explanations would be acceptable if they explain the observations.

16. Quite often two groups of scientists will support opposing theories about some aspect of science. Which one of the following would be the MOST important point to consider in settling such a controversy.

39 A. Both theories give satisfactory explanations for the observations related to the problem, but one theory has more practical applications.
4 B. One group of scientists believe more strongly in their theory.
5 C. One group contains several scientists who have won the Nobel Prize for science.
50 D. Different conclusions are reached when the two theories are applied to certain problems.

17. A scientist shows that he is open-minded when he

15 A. discusses his ideas with other scientists.
58 B. evaluates ideas which do not agree with his

theories.

9 C. agrees with the ideas presented by other scientists.

18 D. asks other scientists to provide experimental evidence to support their arguments.

18. Theories in science are generally accepted when it can be shown that they explain all of the related observations. However, it is possible that exceptions to the theory may exist but are still undiscovered. Which one of the following is the BEST approach to this problem?

42 A. The limits under which the theory has been shown to apply should be carefully stated and the theory should be used within these limits.

16 B. Scientists should provide several theories to explain a given set of observations so that if exceptions to one theory are found, they will have others to rely on.

22 C. Scientists should not accept a theory until they are certain that exceptions to it do not exist.

20 D. When exceptions are discovered, scientists should abandon the theory and look for a new one.

19. Which one of the following is NOT an important reason why scientists often repeat the experiments reported by other scientists?

21 A. A scientist could be so intent on finding a specific answer that he might subconsciously observe only what he wants to see in his experiments.

47 B. This helps to keep scientists careful and honest when making observations and reporting results.

18 C. Other scientists might give a different interpretation to the same observations.

14 D. The first scientist might overlook a significant variable in his experiment.

20. A scientist has a theory for which he needs some evidence. He does experiments and finds that some of the results do not support his theory. When he reports his theory, he omits those results which do NOT fit. In this case, the scientist

A. had a theory which did not have any practical value.

B. considered several possible explanations.

C. made his theory explain part of the experimental results.

D. made the experimental results agree with his theory.

NOTE: Question 20 was omitted from the data analysis because of a typing error in the test copy.

INTENT COMPONENT SUBTEST

21. Scientists have questioned many religious beliefs. Which one of the following best expresses the way you feel concerning this matter?

3 A. When scientific theories question religious beliefs, it is better to keep the religious beliefs.
20 B. I now question all of my religious beliefs since science has cast doubt on some of them.
52 C. I have two separate thought compartments (one for my religious beliefs and one for scientific knowledge).
25 D. I will keep my religious beliefs until scientists prove them to be wrong.

22. Imagine that you have just finished a laboratory investigation. Your measurements all agree except two. Which of the following would you do?

11 A. Include the two odd measurements in your report but omit them from calculations.
3 B. Adjust the two odd measurements to make them agree better with the others.
55 C. Take more measurements.
30 D. Use all the measurements as they are when doing calculations.

23. Consider the following data concerning fluoridation of the public water supply:

> Fluorides help prevent cavities in children's teeth but do not help adult teeth.
> Small amounts of fluorides appear to have no long-term harmful effects.
> The easiest and cheapest way to administer fluorides is through the public water supply.
> The fluoride content of lakes and oceans is increasing as a result of fluorides in the public water supply.
> Fluorides can be put in milk for children.

Which one of the following best describes your point of view after considering the above information?

12 A. You would be against fluoridation.
37 B. You would be uncertain as to which side to support.
39 C. You would be in favor of fluoridation.
12 D. You would lose interest in the problem because the evidence is too indefinite.

24. "Light travels as a stream of particles."
"Light travels as a wave."

If you came across these two statements in two different science books, which of the following would you do?

8 A. Ask your teacher to tell you which statement to accept.
64 B. Check other science books for statements on this topic.
14 C. Assume that scientists are not certain as to how light travels.
15 D. Accept the statement in the newer book.

25. Imagine you are living in a small town on the banks of a river not far from a large industrial city. Your town has just experienced a severe flood for the first time in its history. Some people are saying that it was caused by increased rainfall due to the smog from the nearby industry. Which one of the following best expresses your evaluation of this claim?

35 A. This is a popular opinion for which there is no evidence.
27 B. People are making this claim because of their prejudice against smog.
3 C. This is a valid conclusion based on sufficient evidence.
35 D. This is a popular opinion backed by some evidence.

26. Suppose that you and a friend both did the same experiment to determine whether or not sunlight is required for plants to produce starch. Both of you tested a leaf from a plant that had been left in the dark for two days. Then you both tested a leaf from a plant that had been left in the sunlight. Your friend found starch in both leaves. You found starch only in the leaf from the plant that had been left in the sunlight. Which one of the following would be the most reasonable thing for you to do?

8 A. Accept your own result because text books say that plants in the dark should not produce starch.
87 B. Have both of you repeat the experiment.
1 C. Accept the result obtained by the one of you who knows the most about science.
4 D. Ask your teacher to decide which result should be accepted.

27. Suppose you wanted to determine which types of mosquitoes cause malaria. You obtained three kinds (Types A, B, and C) and examined the digestive tracts of each for malaria parasites. You found some only in

Type B mosquitoes. You concluded from this that malaria is spread by Type B but not by Types A and C mosquitoes. Which of the following describes your conclusion?

1 A. Your conclusion does not agree with the evidence.
18 B. Your conclusion is valid in light of the evidence.
58 C. Your conclusion is justified, but more evidence should be obtained.
23 D. You did not obtain enough evidence to make a conclusion.

28. Some medical researchers say that marijuana does permanent damage to the brain, while others say that it is no more harmful than alcohol. In the light of this information, which of the following would you be inclined to do?

16 A. Not smoke it because it is probably harmful.
9 B. Ignore the evidence that it might be harmful and smoke it if you wanted to.
6 C. Smoke it because it is probably no more harmful than alcohol.
68 D. Put off any decision about smoking it until more definite knowledge is obtained about its effects.

29. "Many people have cycles of mental depression which correspond to the phases of the moon." Which one of the following best represents your reaction to this statement?

60 A. One should be willing to consider the possibility that there may be some truth to superstitions of this nature.
22 B. Scientists could never prove or disprove this idea.
8 C. It is an incorrect idea, but it is useful to many people.
9 D. There seems to be some truth in this statement.

30. Below are a number of points of view regarding the teaching of the theory of evolution in biology class. In your opinion, this theory should be

5 A. omitted from the biology course.
9 B. presented to the class, but its controversial aspects should not be discussed.
49 C. discussed thoroughly in class with all students present.
37 d. discussed openly in class, but those students who do not want to listen should be permitted to leave.

31. Suppose you live near a large industrial plant. You find that the rose bushes in your yard die in a short while, but your lawn remains in perfect condition. You

suspect that the fumes from the industrial plant are the cause. Which one of the following would be the most reasonable course of action for you to take?

81 A. Study the effect of the fumes on healthy rose bushes.
7 B. Stop growing rose bushes.
10 C. Start legal action against the plant for pollution control.
2 D. Move away from the plant.

32. During a class discussion, a friend of yours said, "The questions which are really important to man can never be solved by science." Which one of the following would probably be your reaction to this statement?

1 A. Support him because friends should stick together.
3 B. Not pay any attention to this statement because it is not worth thinking about.
86 C. Ask him to present facts and arguments to support this statement.
11 D. Support him because you believe that the statement is true.

33. Suppose you did a chemistry experiment, but the results were not what you expected. Which one of the following would you do?

4 A. Report the results which were predicted in the chemistry text.
4 B. Copy the results from a friend.
85 C. Report the results that you obtained.
8 D. Report no results and tell the teacher that the experiment failed.

34. A boy goes skating on a pond and breaks through the ice. He is rescued and given a drink of hot chocolate by someone who is sneezing and coughing. A few days later the boy also has a cold. Which one of the following best describes the reason for the boy's cold?

32 A. His cold is due to falling in the cold water and getting wet.
26 B. He got the cold from the person who rescued him.
3 C. He probably had a cold coming before he went skating.
39 D. The reason why people get colds is not yet known for certain.

35. In an experiment, students blew through limewater and noted that it turned milky. From this result, most of them concluded that their bodies give off carbon dioxide. However, one girl wrote in her notebook that since there is carbon dioxide in the air we breathe,

the experiment proved nothing. Which one of the following best describes your evaluation of this situation?

9 A. The students were justified in making their conclusion.
22 B. The girl was justified in doubting the proof.
20 C. Neither side had sufficient grounds for their statements.
50 D. Both sides were partly justified in their statements.

36. "People born when certain stars are becoming more prominent show the influence of these stars in their personalities." People who believe this statement

7 A. probably have a special ability to understand such influences.
18 B. are not critical enough.
41 C. are more open-minded than most people.
34 D. have a disregard for scientific evidence.

37. When evaluating the accuracy of ideas in science texts, which one ot the following is the most important?

31 A. How recently the book was published.
11 B. Whether or not the author is a scientist.
18 C. The extent to which the ideas have been simplified.
39 D. How recently the ideas were first presented.

38. If you came across a scientific idea which goes against your common sense, which one of the following would you be inclined to do?

6 A. Disregard the scientific idea because it is better to rely on common sense.
6 B. Disregard common sense because it is not as reliable as scientific study.
56 C. Do an experiment to see whether or not the common sense is superior to the scientific idea.
33 D. Try to produce a compromise between the scientific idea and common sense.

39. Suppose you had worked several days on a chemistry experiment. You then accidentally added some sodium nitrate solution when you should have added silver nitrate. Which one of the following courses of action would you take?

56 A. Start over again as soon as you realize your mistake.
8 B. Continue with the experiment but if it doesn't turn out the way it should, start over.
34 C. Continue the experiment to see if the mistake makes

any difference.

2 D. As soon as you realize your mistake, add some silver nitrate solution and continue with the experiment.

40. A missionary reported that the root of a plant much like the Rauwolfia plant had been used by an African witch doctor to cure him of a serious illness. Recent medical reports show that reserpine, a drug effective in lowering blood pressure, is extracted from Rauwolfia. Which one of the following is the most reasonable conclusion that can be drawn from the above discussion?

15 A. Since the witch doctor probably did not know anything about modern drugs, he did not have a scientific reason for using the roots.

6 B. The plant was probably not helpful because the missionary had no way of knowing what caused him to get better.

34 C. The plant may have been helpful since the missionary recovered after the witch doctor's treatment.

45 D. The plant probably was helpful because the Rauwolfia plant contains reserpine.

APPENDIX C

TEST OF LIKERT ITEMS

This part of the test consists of 25 statements. You are asked to indicate whether you agree with each of these statements. Mark the answer sheet according to the following key:

Mark A if you STRONGLY AGREE with the statement.
Mark B if you PARTLY AGREE with the statement.
Mark C if you PARTLY DISAGREE with the statement.
Mark D if you STRONGLY DISAGREE with the statement
IGNORE column E on the answer sheet.

Example: Answer Sheet

200. People depend on plants and animals for food. 200. A1 B2 C3 D4 E5

Since A is marked, this would indicate that you strongly agree with this statement.

Some of the Likert items in this list were taken from the science attitude scales discussed in Chapter II. The references for these items are included in this appendix. The remainder of the statements were written for use in the present study. The numbers to the left of each statement are the percentages of students out of 156 who chose the indicated response. The response which was assigned a 4 in the scoring of the test is underlined for each statement.

SA	PA	PD	SD		
54	34	8	4	1.	When a scientist is shown enough evidence that one of his ideas is a poor one, he should change it. (Moore and Sutman, 1970)
93	5	2	0	2.	It is very important for a scientist to report exactly what he observes.
4	11	13	72	3.	Investigation of the possibility of creating life in the laboratory is an invasion of science into areas where it does not belong.
74	22	3	1	4.	Scientists sometimes repeat experiments done by other people to check their results.
42	45	9	4	5.	Scientists should criticize each others work. (Moore and Sutman, 1970)
83	17	0	1	6.	It is more important to get along with people than to make them angry by trying to convince them that they are wrong.
1	6	21	72	7.	Once a good theory is developed, scientists should not question it.
6	3	11	80	8.	When reporting his results, a scientist should omit those which do not support his theory.
4	38	36	22	9.	If a few scientists have evidence which appears to contradict a current scientific theory, then the theory is probably wrong.
32	58	7	3	10.	Many ideas which scientists find to be useful may not be entirely correct.
51	36	10	4	11.	It is necessary to question periodically the basic truths of science.
7	31	27	35	12.	The skeptism of the scientist should be

limited to his work. (Schwirian, 1968)

3 4 10 83 13. It is useless to listen to a new idea unless everybody agrees with it. (Moore and Sutman, 1970)

4 33 33 29 14. When something is explained well, there is no reason to look for another explanation. (Moore and Sutman, 1970)

2 3 26 64 15. Scientific findings should not be made public if they will create controversy. (Schwirian, 1968)

6 12 36 46 16. When the findings or theories of science conflict with religious belief, it is better to accept the religious belief. (Schwirian, 1968)

13 24 29 34 17. When some new facts are discovered which are not explained by an existing theory, the unexplained facts may be revised or ignored. (Kimball, 1968)

75 24 1 1 18. A person should not make up his mind until he has collected as many facts as possible.

10 25 37 28 19. Before accepting a new theory, a scientist would want to know how well it explains the facts.

64 22 10 3 20. Scientists should be free to explore all aspects of man's life and the universe about him. (Schwirian, 1968)

4 15 37 44 21. Once a person makes up his mind he should be reluctant to change it.

51 33 12 4 22. Religious leaders should take into account the ideas which scientists explore and the theories they produce.

15 28 31 26 23. When making decisions about drinking alcohol and smoking, personal preferences are more important than the results of scientific studies.

79 19 1 2 24. It's important to try and figure out why an experiment which you have done did not turn out the way the lab manual said it should.

4 23 34 38 25. It is alright for a student to say he verified a scientific law, even if his experimental results were not too good.

APPENDIX D

Instructions to Panel For Rating Behaviors

Panel member:

The affective domain in science has been defined in terms of affective attributes which scientists exhibit in their work and in their relationships with their peers (Nay and Crocker, 1970). I feel that many of these attributes can also be observed in students as they do science work at school. The identification of the presence or absence of these attributes in students can serve as a basis for evaluation of affective objectives in science education.

I am constructing test items to identify the following attributes; critical mindedness, questioning attitude, suspended jusdement, respect for evidence, honesty, objectivity, willingness to change opinions, open-mindedness, and disciplined thinking. As a first step, these attributes will be defined in behavioral terms. I would greatly appreciate your assistance in this task.

I have listed a number of behaviors for each attribute. Please indicate your opinion regarding the extent to which each behavior defines the attribute under which it is listed, that is, to what extent would the presence of this behavior indicate that the individual possesses that attribute? Check the appropriate square according to the following key:

I. If you feel that the behavior is trivial or that the behavior is not related to the given attribute.

II. If you feel that the behavior is an important defining characteristic of the attribute

III. If you feel that the behavior is a very important defining characteristic of the attribute.

These three categories will be weighted on the following scale:

I-0, II-1, III-2

A student demonstrates critical mindedness when he:

I	II	III	TOTAL[1]	
1	8	2	12	-re-evaluates previous solutions to problems
4	4	3	10	-subjects his own ideas to evaluations by others
4	5	2	9	-makes his own observations
0	1	10	21*	-looks for inconsistencies in statements and conclusions
1	6	4	14*	-consults a number of authorities when seeking information
5	4	2	8	-searches for new methods of investigating a problem
0	3	8	19*	-looks for empirical evidence to support or contradict explanations
3	5	3	11	-repeats a procedure several times to compare results
2	7	2	11	-examines many sides of a problem and considers several possible solutions
1	6	4	14*	-asks many questions starting what, where, why, when and how
2	6	3	12	-insists on hearing more than one opinion on a topic
1	2	8	18*	-challenges the validity of unsupported statements

A student demonstrates suspended judgment when he:

I	II	III	TOTAL	
2	7	2	11	-repeats procedures several times and compares results
3	6	2	10	-examines many sides of a problem and considers several possible solutions
1	4	6	16*	-generalizes only to the degree justified by available evidence
1	3	8	19*	-collects as much data as possible before drawing conclusions
1	3	7	17*	-recognizes conclusions as being tentative
0	9	2	13*	-consults several authorities (texts, periodicals, people) before drawing conclusions
3	5	3	11	-recognizes that knowledge is incomplete

---

[1] The first three columns contain the number of panel members out of 11 that responded I, II and III. The total is the following sum Ix0 + IIx1 + IIIx2. The totals are followed by an * for those behaviors which were selected for the final definitions.

A student demonstrates respect for evidence when he:

I	II	III	TOTAL	
0	2	9	20*	-looks for empirical evidence to support or contradict explanations
1	8	2	12	-makes his own observations
1	7	4	15*	-collects as much data as possible before drawing conclusions
5	4	2	8	-collects data to determine the degree of reliability of common superstitions
0	6	5	16*	-demands that explanations fit the facts
0	2	9	20*	-demands supportive evidence for unsubstantiated statements
0	5	6	17*	-supplies empirical evidence to support his statements

A student demonstrates honesty when he:

I	II	III	TOTAL	
0	2	9	20*	-reports observations even when they contradict his hypotheses
0	6	5	16*	-acknowledges work done by others
1	4	6	16*	-considers all available information when forming generalizations and drawing conclusions
5	3	3	9	-states the basic assumptions inherent in solving a problem
3	6	2	10	-reports many sides of an argument
8	3	0	3	-offers constructive criticism of other peoples' work

A student demonstrates objectivity when he:

I	II	III	TOTAL	
2	7	2	11	-bases his conclusions upon evidence from a variety of sources
0	0	11	22*	-considers all available data (not only that portion which supports his prior hypotheses)
3	3	5	13	-listens to several opinions on a topic
2	8	1	10	-reads several sources expressing various aspects of a given topic
5	4	2	8	-makes statements only when they can be substantiated
0	3	8	19*	-reports observations even when they contradict his hypotheses
0	7	4	15*	-considers and evaluates ideas presented by others
1	5	5	15*	-examines many sides of a problem and considers several possible solutions
0	7	4	15*	-considers both pros and cons when evaluating a situation

A student demonstrates willingness to change opinions when he:

I	II	III	TOTAL	
1	3	8	19*	-recognizes conclusions as being tentative
0	8	3	14*	-recognizes that knowledge is incomplete
0	7	4	15*	-considers and evaluates ideas presented by others
4	4	3	10	-evaluates reports of new theories
2	7	2	11	-seeks and considers new evidence
0	8	3	14*	-evaluates evidence which contradict his hypotheses
0	2	9	20*	-alters his hypotheses when necessary to accommodate empirical data

A student demonstrates open-mindedness when he:

I	II	III	TOTAL	
0	7	4	15*	-considers and evaluates ideas presented by others
0	7	4	15*	-evaluates evidence which contradicts his hypotheses
3	4	4	12	-alters his hypotheses when necessary to accommodate empirical data
4	6	1	8	-reads several sources expressing various aspects of a given topic
3	5	3	11	-is willing to listen to several opinions on a given topic
3	8	0	8	-evaluates reports of new theories
1	5	5	15*	-considers several possible options when investigating a problem
0	5	6	17*	-considers both pros and cons when evaluating a situation

A student demonstrates disciplined thinking when he:

I	II	III	TOTAL	
0	7	4	15	-organizes data for the purpose of forming generalizations
1	8	2	12	-organizes new knowledge in terms of concepts, models and theories
3	4	4	12	-makes predictions of possible outcomes or solutions when faced with a problem
0	2	9	20	-generalizes only to the degree justified by available evidence
0	6	5	16	-distinguishes relevant from non-relevant data when attempting to solve a problem
4	5	2	9	-presents reasons to support choices which must be made
0	2	9	20	-carries a given line of reasoning to a logical conclusion

A student demonstrates questioning attitude when he:

I	II	III	TOTAL	
2	7	2	11	-re-evaluates previous solutions to problems
3	3	5	13	-subjects his own ideas and data to evaluation by others
8	1	2	5	-makes his own observations
0	4	7	18*	-looks for inconsistencies in statements and conclusions
0	9	2	13*	-consults a number of authorities when seeking information
3	5	3	11	-searches for new methods of investigating a problem
0	5	6	17*	-looks for empirical evidence to support or contradict explanations
3	6	2	10	-repeats a procedure several times and compares results
2	5	4	13	-examines many sides of a problem and considers several possible solutions
2	2	7	16*	-asks many questions starting who, what, where, why, when and how
1	9	1	11	-insists on hearing more than one opinion on a topic
1	2	8	18*	-challenges the validity of unsupported statements

APPENDIX E

Instructions To The Teachers For Student Ratings

I have developed test items to identify the following characteristics: critical mindedness (questioning attitude), suspended judgement (restraint), respect for evidence (reliance on fact), honesty, objectivity (open-mindedness), and willingness to change opinions. Some of the test items require the students to recognize how these characteristics influence scientists in their work. The other test items ask the students to indicate the extent to which they would exhibit these characteristics in various situations. The test items are based on the following definitions:

The student demonstrates critical mindedness when he:

Looks for inconsistencies in statements and conclusions.
Looks for evidence to support or contradict explanations.
Challenges the validity of unsupported statements.
Consults a number of authorities when seeking information.
Asks many questions beginning why, who, what, where, and how.

The student demonstrates suspended judgement when he:

Generalizes only to the degree justified by available evidence.
Collects as much data as possible before drawing conclusions.
Recognizes conclusions as being tentative.
Consults several authorities (texts, periodicals, people) before drawing conclusions.

The student demonstrates respect for evidence when he:

Looks for evidence to support or contradict explanations.
Demands supportive evidence for unsubstantiated statements.
Supplies evidence to support his statements.
Demands that explanations fit the facts.
Collects as much data as possible before drawing conclusions.

The student demonstrates honesty when he:

Reports observations even when they contradict his hypotheses.
Acknowledges work done by others.
Considers all available information when forming generalizations and drawing conclusions.

The student demonstrates objectivity when he:

Considers all available data (not only that portion which supports his prior hypotheses).
Considers both pros and cons when evaluating a situation.
Examines many sides of a problem and considers several possible solutions.
Considers and evaluates ideas presented by others.

The student demonstrates willingness to change opinions when he:

Recognizes conclusions as being tentative.
Recognizes that knowledge is incomplete.
Alters his hypotheses when necessary to accommodate data.
Evaluates evidence which contradict his hypotheses.
Considers and evaluates ideas presented by others.

Please reread these definitions and try to form an overall picture of what I am trying to measure with my test. The purpose of the student rating which I will be asking you to do is to identify those students which tend to exhibit these characteristics to a high degree and those students which do not exhibit these characteristics. You will be asked to make one rating for each student, that is, to rate the students on the total set of behaviors listed above, not on each separate characteristic. You will be asked to rate each student on a four point scale. In your opinion, student A

1. ______________	2. _________	3. _________	4. ______________
does not exhibit these behaviors to any noticeable degree	increasing tendency to exhibit these behaviors		frequently exhibits these behaviors

The following points should be considered when giving your rating:

Your observations of the student's behavior in science class.
The student's participation in science clubs or related extra-curricular activities.
The student's use of the library facilities.

Do not use the student's marks on science achievement tests as a criterion for giving your rating.

APPENDIX F

TETRACHORIC CORRELATIONS FOR 39 TEST ITEMS FROM TOSA FOR 307 STUDENTS

ITEM	1	2	3	4	5	6	7	8	9	10	11	12	13
1	100												
2	-18	100											
3	27	04	100										
4	-15	11	-05	100									
5	10	10	07	01	100								
6	-01	14	13	-06	34	100							
7	16	26	27	02	27	27	100						
8	18	02	29	-18	05	04	-03	100					
9	12	04	-01	-08	02	03	18	08	100				
10	14	26	20	-01	06	15	34	12	36	100			
11	-14	-14	03	12	05	-08	-12	08	13	11	100		
12	12	11	07	04	-03	25	21	13	03	22	-02	100	
13	13	20	01	13	13	05	15	02	07	22	-08	14	100
14	07	-14	14	10	09	01	06	04	01	28	-08	03	09
15	15	-19	10	-12	10	13	06	28	-05	20	-10	15	-04
16	-03	-07	20	-10	05	-03	01	-01	29	14	01	-09	13
17	17	08	25	-06	12	25	20	11	19	27	06	10	27
18	03	-01	14	12	26	06	15	03	11	20	-03	06	20
19	-17	-14	-21	28	-17	-10	-30	-09	14	-30	05	-10	-19
21	02	01	17	-01	01	-01	06	07	-08	-08	-16	05	07
22	-10	05	05	18	-36	23	-11	-13	-36	-36	-29	-04	03
23	13	15	04	03	10	10	-03	-04	-01	11	01	-02	07
24	-05	-04	-11	04	-01	18	10	-13	08	-01	14	13	02
25	-01	12	-02	09	06	07	19	-21	11	06	04	03	-07
26	15	17	01	28	31	43	11	-07	42	43	13	10	35
27	02	09	14	04	12	01	17	-02	-03	09	-10	-01	30
28	18	17	-10	11	04	12	08	-02	10	17	-04	01	02
29	23	15	09	09	17	07	17	-04	15	21	-03	-01	28
30	08	07	04	14	01	05	03	03	16	25	-01	15	10
31	03	04	01	20	-04	10	10	-18	03	10	16	10	-18
32	29	-17	-08	-03	21	29	08	02	23	30	06	16	01
33	-21	10	04	12	06	14	11	14	55	43	30	39	-15
34	07	12	11	15	-01	-01	13	07	10	14	16	03	15
35	-17	09	-01	-02	12	22	16	05	-06	23	-07	01	-03
36	-06	14	20	-09	13	02	20	-04	-12	22	-04	-05	-07
37	01	10	-02	-02	-08	08	-04	21	-04	-05	10	-05	-11
38	-11	-01	-10	13	-01	-04	03	-22	16	11	05	-02	-09
39	21	-02	-17	06	18	12	-04	07	-07	10	08	-07	07
40	-02	10	-09	-11	-05	-04	-06	11	-11	13	-07	05	-20

The item numbers in the correlation matrix correspond to the item numbers in Appendix B. Item 20 has been omitted. The correlation coeficients have been multiplied by a factor of 100.

ITEM	14	15	16	17	18	19	21	22	23	24	25	26	27
14	100												
15	02	100											
16	03	-06	100										
17	10	26	14	100									
18	15	27	06	19	100								
19	05	14	11	-13	02	100							
21	13	05	06	05	-03	-15	100						
22	01	-13	-08	-10	-10	14	-29	100					
23	-08	01	02	07	-01	07	-13	-16	100				
24	-06	02	-03	07	05	-03	-13	19	-13	100			
25	-08	-03	-06	-05	-07	12	-05	-06	-03	-03	100		
26	-07	02	12	15	25	-11	-04	31	25	30	-02	100	
27	-04	-11	13	04	16	-03	18	11	-01	12	02	11	100
28	-10	10	-06	07	12	-17	15	32	-08	-01	-01	04	10
29	01	-09	06	-16	-01	-16	09	-10	31	-17	-01	21	20
30	14	-08	15	19	19	-11	28	-24	16	05	15	17	08
31	-02	-08	13	-08	-09	-17	01	24	10	04	01	22	-03
32	03	-18	-08	-14	-01	-17	19	20	-08	15	17	46	18
33	-09	25	16	28	39	-02	20	-04	-15	18	-10	45	-02
34	-10	18	13	-07	04	01	22	-07	02	-01	07	22	06
35	04	14	-11	19	14	02	08	14	10	-01	-02	21	05
36	03	07	13	14	04	-06	03	22	23	-01	-01	13	10
37	-13	05	-09	01	03	12	04	13	04	21	13	07	-07
38	-08	-11	10	-04	-10	02	-16	04	-02	09	01	31	03
39	05	-02	01	-01	-11	02	-06	-36	-10	17	13	16	-17
40	-07	-08	-06	-09	-05	03	-07	07	08	-05	04	-03	09

ITEM	28	29	30	31	32	33	34	35	36	37	38	39	40
28	100												
29	17	100											
30	09	-04	100										
31	38	23	14	100									
32	41	15	08	41	100								
33	23	-05	10	52	26	100							
34	22	23	04	13	19	26	100						
35	18	15	-08	05	26	06	-02	100					
36	-18	32	-09	-03	-19	-28	-09	04	100				
37	-04	10	07	04	11	23	-02	03	14	100			
38	10	-01	08	16	27	16	10	-18	07	-14	100		
39	-10	-02	-07	02	17	-07	06	-01	-10	18	06	100	
40	19	-01	08	01	06	-07	-15	16	04	22	-11	11	100

APPENDIX G

INTERCORRELATIONS FOR THE NINE FACTORS

FACTOR	1	2	3	4	5	6	7	8	9
1	100								
2	18	100							
3	-05	-01	100						
4	01	-02	02	100					
5	-06	-10	14	-16	100				
6	01	-01	01	01	05	100			
7	05	02	-11	02	-15	01	100		
8	05	-01	-01	-02	04	07	-02	100	
9	-08	-11	-01	-08	11	07	-07	-03	100

The correlation coefficients have been multiplied by a factor of 100.